INVENTORY OF FEDERAL AND STATE WATER DATA

A WHITE PAPER ON BUILDING A DATA INVENTORY

JULY 2019

Duke
NICHOLAS INSTITUTE
FOR ENVIRONMENTAL POLICY SOLUTIONS

Internet of Water

Inventory of Federal and State Water Data

Last Updated: July 23, 2019
Authors: Lauren Patterson and Nicole Abib

Introduction

The Internet of Water (IoW) is a bold vision for how to improve our nation's water data infrastructure — transforming water management by making water data more discoverable, accessible, and usable. The water challenges of the 21st century are increasingly complex and interconnected. The depletion of multi-scale aquifers, basin-scale flooding, or the widespread accumulation of nutrients leading to dead zones span across multiple jurisdictions and are the product of decades of individual and collective decisions. Data will not fix these problems, but they can help us understand and better manage water resources in real-time. Unfortunately, while water data have been collected by federal, state, and local agencies for decades, much of the data remain difficult to find, to access, or to integrate across, or even within, a single jurisdiction. The purpose of the IoW is to provide a sustained effort to accelerate ongoing — and catalyze new — efforts to make public data more discoverable, accessible, and usable.

An important first step for the IoW is to understand what publicly collected water data already exist, particularly data related to a water budget. What is the quantity (how much water is there), quality (how clean is the water), and use (what purposes is water serving). The IoW has begun to inventory public governments to identify (1) who is collecting water data, (2) the purpose of collection, (3) types of water data collected, and (4) how FAIR (findable, accessible, interoperable, and reusable) the data are for secondary data users. As such, we sought to begin the process of understanding how FAIR water data currently are and some best practices for making water data more FAIR. The inventory began with the federal government, as well as three state governments: California, North Carolina, and Texas. The data and results of the inventory can be explored here: http://internetofwater.org/public-data-inventory/.

The IoW data inventory is based on an outsider's perspective of data FAIRness. The scores generated are solely based around data openness and do not take into consideration that these data platforms (platform simply refers to the place online where data are found and accessed and is used interchangeably with website, data access point, etc.) were rarely designed to be FAIR. Often, platforms are designed to meet the needs of a specific community and may be meeting those needs well. The data in the inventory may also become outdated quickly. The data in this initial inventory are accurate as of November 2018.

Methods

The data inventory was built by locating the organizational chart for the executive branch of federal and state governments. Since each government used different terms to describe their organizational hierarchy, we use the term "entity." **Entity** refers to a component within the executive branch collecting water data. Entities may refer to a department, division, agency, river basin authority, program, etc.

The analysts went to the website for each entity and searched for water data. The website providing the data is referred to as a "platform." **Platform** simply refers to the online location where data are found and accessed. Platforms may be websites, catalogs, maps, etc. Each platform was assessed based on a suite of FAIR metrics. Metrics were recorded and categorized based on the best judgment of the analysts based on what was publicly available at the time of the inventory. States and federal agencies are actively updating platforms, which could result in metrics being out-of-date. The inventory is a living document, welcoming feedback from state and federal agencies for improvement. The metadata, templates, and process for creating a data inventory are available for download at http://internetofwater.org/public-data-inventory/.

Entities Collecting Water Data

Each inventory began by locating the organizational chart for the executive branch of the selected government. The organizational charts served as the starting point for locating which entities had water as a part of their mission or provided water data on their websites. Data collected included entity name, mission, and the URL. The mission and website content shaped how we cataloged the purpose behind collecting water data. The categories of data purpose were:

- **Operational.** Data are used to inform day-to-day operations.
- **Broad Scale Decision-Making**. Data are used to inform policy, investment, management and other types of decision support systems that span a large geographic region or period of time.
- **Regulatory**. Data are collected by mandate, typically to ensure compliance for public or environmental health.
- **Research and Innovation.** Data are used to advance knowledge and innovate.

An entity may collect data for multiple purposes. The data are visualized as a hierarchical network and allows users to highlight and filter the network based on entity and data purpose (Figure 1).

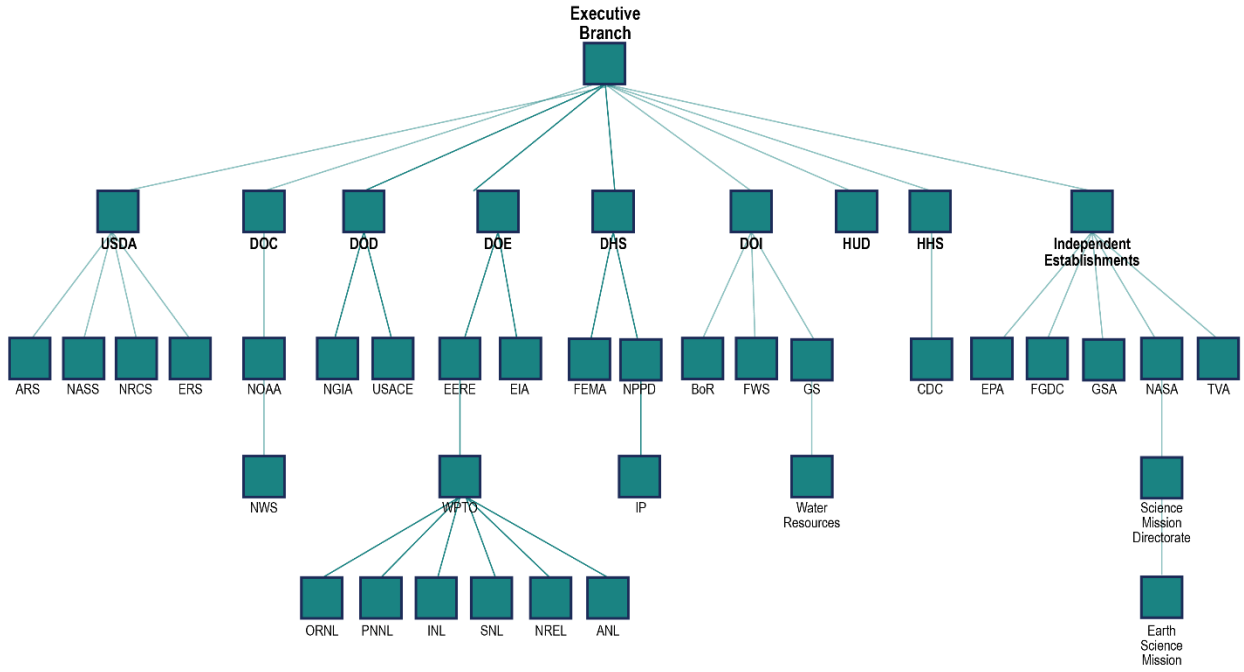

Figure 1. Organizational structure of federal government entities collecting water data.

Water Data Platforms

Entity websites were searched to locate where data were found and accessed. The types of data provided were cataloged at each platform, along with attributes related to FAIRness. FAIRness metrics were designed to enable comparison of data openness within and across inventories.

Water Data Types

Public agencies collect a wide variety of water data. The types of water data collected were cataloged by the basic components of a water budget: quantity, quality, and use. Infrastructure, whether natural or built, allows water to move between systems and became an additional component of the water budget. Water budget categories were further subdivided into additional categories (Table 1).

Table 1: Categories and subcategories of water data types.

Category	Subcategory	Types of data included
Infrastructure	Built	Dams, levees, diversion locations, utility locations, wells, etc.
	Natural	NHD, NHDplus, watershed boundaries, aquifer boundaries, wetlands, etc.
Water Quality	Quality	All constituent measurements for surface water and groundwater quality, including sedimentation
	Regulatory	303d Assessment, Facility permits, compliance, violations, and enforcement actions, TMDLs, regulated and unregulated contaminant occurrence, consumer confidence reports, etc.
Water Quantity	ET	Evaporation, evapotranspiration
	Extreme Events	Flood boundary maps and hazard layers, Flood Insurance Rate Maps, flood risk maps, Hazus, Flood stage and frequency, Palmer drought indices, National Drought Map, etc.
	Glacial and Snow	Glacier extent and speed, ice sheet elevation, temperature, and thickness; snow depth, snow water equivalent, snowfall, etc.
	Groundwater	Depth to groundwater, flow, storage, etc.
	Meteorology	Air/Atmospheric temperature, relative and specific humidity, wind speed and direction, etc.
	Precipitation	Amount, frequency, intensity
	Reservoir	Stage, inflow, volume allocation, storage, outflow, surface area, lake evaporation, water diversions, etc.
	Soil	Soil moisture and temperature, freeze/thaw/melt status
	Surface Water	Stream stage and discharge, streambed elevation, rating curves, stream density and length, etc.
Water use	Hydropower	Power generation; plant capacity, ownership, and type; potential capacity from non-powered dams and undeveloped streams
	Irrigation	Area irrigated and source; canal discharge, diversion, and stage; irrigation water applied, number of operations, irrigation supply condition, etc.
	Use	Water pumped, water use by sector, water supply source, etc.
	Utilities	Quantity, quality, and use of water supplied by public municipalities as well as domestic water usage.
	Management Plans	Includes state and regional watershed, aquifer, and water supply management plans.

Network Creation

A network was created in JavaScript using vis.js to show the relationships between data platforms and the types of water data provided. Edges are colored based on data category and visually illustrates which entities collect similar data types, which data types are rarely collected, which platforms provide multiple types of data, and so on (Figure 2). The size of the data node reflects the number of platforms providing that type of data. Platforms nodes collecting similar types of data cluster while platforms providing a single type of data, particularly if few other platforms provide that type of data, will be located on the periphery of the network. Platforms not providing data, or data types not provided by public agencies are isolated nodes. Users can highlight nodes by platform, water data type, and characteristics of discoverability, accessibility, and usability.

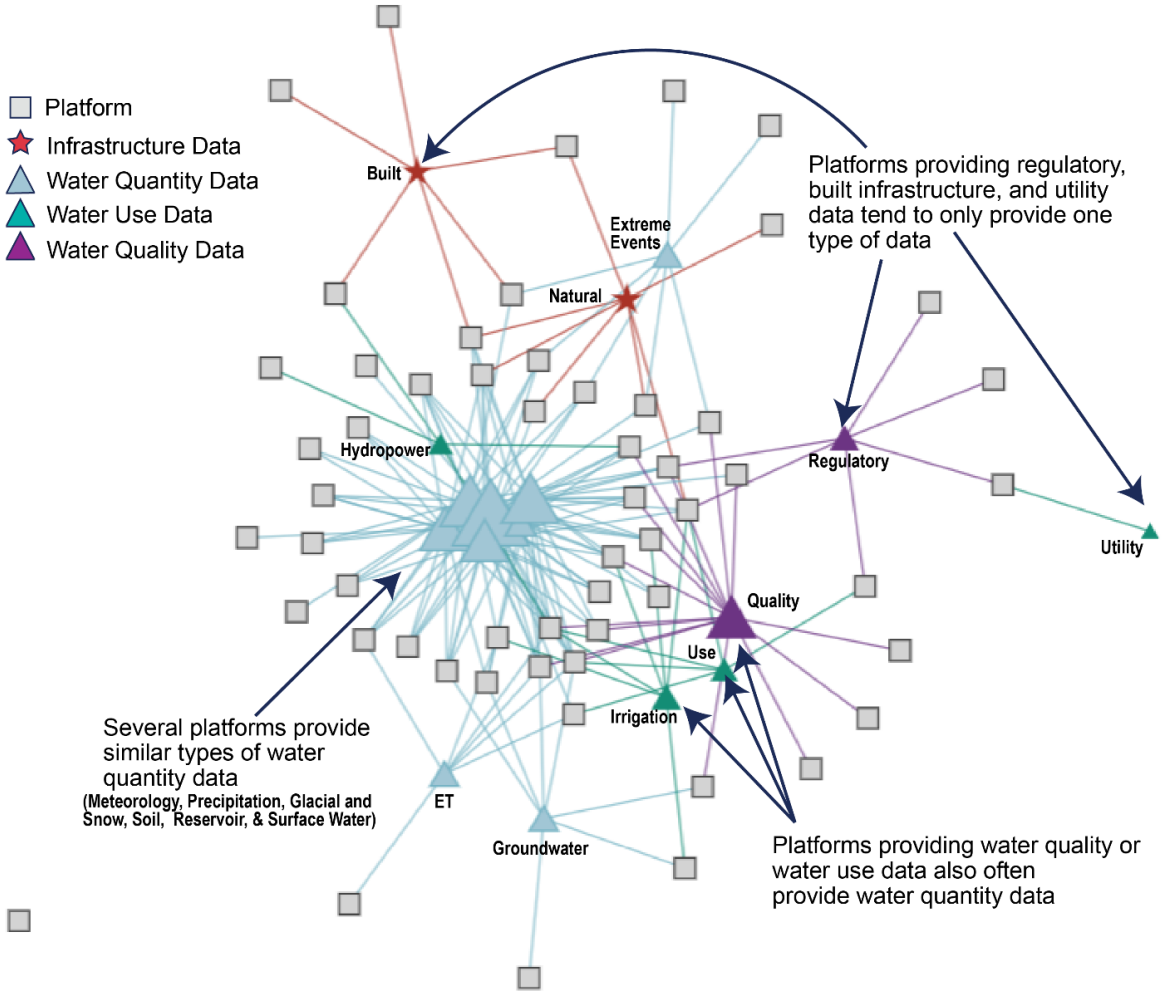

Figure 2. The network shows the relationship between federal data platforms and types of water data.

Summary of Water Data Collected

A heat map summarizes which platforms provide similar types of data and the variety of data types collected (Figure 3). For instance, the federal government heat map shows that the Environmental Protection Agency (EPA) collects the majority of regulatory data while the National Aeronautics and Space Administration (NASA) satellites and Department of Commerce, which houses the National Oceanic and Atmospheric Administration (NOAA), provide the majority of the meteorological and precipitation data. The Bureau of Reclamation's (BoR) Water Information System provides the greatest variety of data types (10), followed by the Geological Survey's (USGS) National Water Information System (9) and Water Alert (8) platforms. Precipitation is most frequently collected (n=27 platforms hosted by 10 federal entities). Entities collecting precipitation data typically collected other types of meteorological data too (n = 20 platforms). Eleven federal entities provided surface water data through 19 platforms.

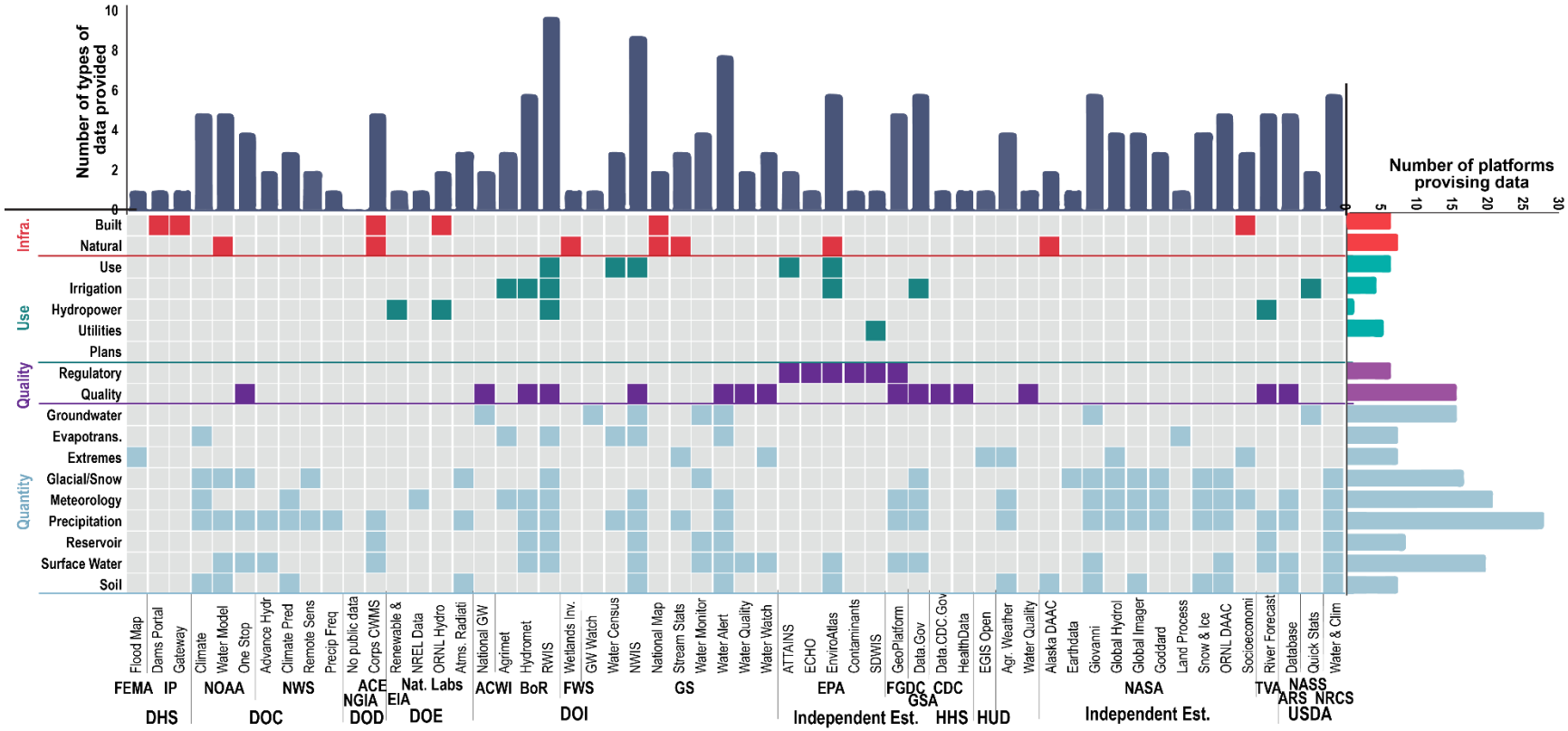

Figure 3: Heat map of water data collected by federal entities. The top chart shows the number of types of data collected by each entity, while the right-hand chart shows the number of entities collecting a particular type of data.

Openness Scorecard

A suite of FAIR metrics were collected for each platform with two metrics related to findability, two metrics related to accessibility, and eight metrics related to interoperability. No metrics were collected on reusability because it was assumed that open, public data are inherently reusable. Scores for each metric were based on an outside user's perspective. These metrics in this inventory are simpler and differ from FAIR metrics recently developed by the FAIR metrics group. Those creating a data inventory may choose to use either suite of metrics.

Findability Metrics

Findability refers to how discoverable data are for both humans and computers. The analysts assessed the ease and the method of finding data. Note that there may be many instances were data were not discoverable and were not accounted for in this inventory.

Ease of Finding Data

The ease of finding data ranged from none (in instances where the website indicated data were collected but the analyst could not find the data), low, medium, and high (Table 2). If a platform was missing from this inventory, the data were not easy to find or the platform was created after the inventory was completed. Scores ranged from zero (no findability) to three (high findability).

Table 2. Categories and scores used to assess the ease of finding data.

Category	Definition	Score
None	Data collected but not found.	0
Low	Data were difficult to find within an entities website, were scattered across multiple URLs, and were not searchable.	1
Medium	Data were relatively easy to find, were found at a single or few URLs, and had poor or inconsistent search index.	2
High	Data were easy to find, were found at the same URL, and had good search indices and well-defined keywords.	3

Method for Finding Data

The methods for finding data generally fell into one of four categories: unknown, point and click on a website, catalog, and/or a map (Table 3). Some platforms provided a combination of these methods, such as searching for data on a map and filtering results through a catalog. The mean score was used when a platform employed multiple methods.

Table 3. Categories and scores used to assess the method for finding data.

Category	Definition	Score
Unknown	Method was unknown without seeking permission, registration, or an additional step.	0
Websites	Data found by navigating through individual URL links.	1
Catalog	Connects data sources to a searchable, descriptive directory so the user can find and explore the data no matter where it is located.	2
Map	Data found based on a location (e.g., watershed, aquifer, or county).	2

Accessibility Metrics

Accessibility refers to how easy it was to obtain the data. Accessibility consists of two components: (1) does the user have permission to access the data? and (2) how does the user retrieve the data?

Ease of Accessing Data

The ease of obtaining the data were scored in terms of general public access. In some instances, the public does not have access to the data. In other instances, the public might need to request permission, receive special training, and/or register for access (Table 4). Those required to register for access were given the same score as platforms providing full access when registration automatically guaranteed access. Registration in this manner was not assigned a lower score because it allows the data provider to understand who uses their data, and potentially how the data are being used to create value. A lower score was assigned when training or software are required, and/or permission was not guaranteed.

Table 4. Categories and scores used to assess the ease of obtaining data.

Category	Definition	Score
No Access	Platform indicated data exist, but did not provide any information or means to access the data.	0
Permission Required	User must request permission to access data. Permission was not guaranteed.	1
Software Required	User must download or obtain special software to access data or undergo training prior to use.	1
Registration Required	User must register to access data with access automatically granted.	2
Full Access	User has full access to data.	2

Method for Accessing Data

The method for obtaining data refers to how data were downloaded and whether the data could be downloaded individually, filtered, or in full (Table 5). Copy & paste and individual exports allowed the user to access the data with significant effort. In contrast, web/API services enabled users to seamlessly access large volumes of data. A platform may provide multiple methods to access data, potentially

meeting the needs for a spectrum of users and skill levels (non-expert and computer programmers). The score was the sum of individual method scores normalized by the number of methods provided.

Table 5. Categories and scores used to assess the method(s) for obtaining data.

Category	Definition	Score
None or Unknown	The method for accessing data was not evident without gaining permission or registering, or there was no access.	0
Copy & Paste	Data were not downloadable but must be copied and pasted from the website.	1
Link to Source	Provided a link to the data producer, not the data. If the link was directly to the data, the method for download was recorded. If the link was to a general website, no download method was recorded.	1
Individual Export	Data can be downloaded but only by one site, attribute, etc. at a time.	2
Batch or Full Export	Data can be filtered such that all selected data may be downloaded simultaneously, within download constraints of the system.	3
FTP	Data were accessed through File Transfer Protocols (FTP).	4
Web/API services	Data were accessed via web services that provide a standardized way to share data and information between devices in a network.	5

Interoperability Metrics

Interoperability, or usability, occurs when systems and services allow for the creation, exchange, and consumption of data with clear, shared expectations on the content, context, and meaning of the data. Some aspects of interoperability are universal: (1) are the data machine-readable and (2) are metadata provided and are those metadata adequate? Other aspects are dependent on the purpose of the data, such as (3) timeliness of updates and (4) length of record. For example, a utility operator might require 15-minute water quality data over the last 24 hours to make decisions, while a state might need decades of monthly data to assess long-term water infrastructure needs. Here, we assume that higher frequencies of data and longer periods of record are ideal; however, we acknowledge that some data purposes would be served well with data provided annually or with only five days of data.

File Formats

File formats were documented and categorized by machine readability (Table 6). Machine-readable formats are necessary for data to be interoperable. Machine-readable formats were categorized by whether they required proprietary software (limits usability) or open software. Sometimes multiple file formats were available (for instance a shapefile and a geojson option, or a csv and Excel option). The maximum score was assigned when multiple file formats were available to represents the maximum potential.

Table 6. Categories and scores based on file formats.

Category	Definition	Score
Not machine readable	Data cannot be downloaded and directly used. File formats include images (jpeg; tiff; png) and pdfs	0
May be machine readable	Data were downloaded in a .zip file, requiring additional effort to access	1
Machine readable with proprietary software	Data were machine-readable but required access to proprietary software to use. File formats include xlsx, docx, shp, etc.	3
Machine readable with open software	Data were machine-readable and include csv, txt, xml, json, geojson, netCDF, etc.	5

Metadata

Metadata are incredibly important because they allow the user to have confidence in the data. Standardized terms, data, and metadata allow for the integration of similar data from multiple sources. There was a wide spectrum in the quality of metadata provided, requiring several metrics to capture format, types, standards, and the presence of a data glossary.

Metadata Format

Metadata come in a variety of formats. Similar to the file format of data, metadata may be provided in machine-readable formats that make it easier for users to assess (Table 7). Sometimes the metadata are part of a pdf or standards of operating procedure that are hard to locate or find the desired information. In other instances, metadata were provided online, embedded within a data file, or download (within a zip file).

Table 7. Categories and scores based metadata format.

Category	Definition	Score
No or Broken Link	No metadata were provided and/or the link was broken.	0
Unknown	Data could not be accessed. It is unknown if metadata are provided.	1
Not machine-readable	Metadata was not in a machine-readable format. It may be located within a pdf.	1
Machine-readable with proprietary software	Metadata were in a machine-readable format that requires proprietary software to access. File formats include excel and word.	2
Machine-readable with open software	Metadata were in a machine-readable format (e.g., xml). Sometimes the metadata were part of the data (e.g, USGS stream gauge data provides metadata within the data download by point and click and separately by web services).	3

Data Definitions

Definitions vary between agencies and glossaries/dictionaries that provides explicit descriptions for each column is necessary to ensure data are interpreted and used correctly. For instance, is water use

referring to withdrawals or consumptive use? Here, we simply note whether a dictionary or glossary was available for columns within datasets (Table 8). The presence of data glossaries can help prevent future misuse from misinterpreting the data.

Table 8. Categories and scores based on the presence of data definitions.

Category	Definition	Score
No or Unknown	No data glossaries found or data were not publicly accessible; it is unknown if a data glossary exists.	0
Some	Some data had glossaries.	1
Yes	Data glossaries were available for all data.	2

Metadata Attributes

There are several types of metadata, but for simplicity the analysts documented the presence of administrative, descriptive, and structural metadata. Administrative metadata provides information about the data, such as technical information, units, quality control, and so on. Descriptive metadata enables data to be discovered and identifies the data producers, including information such as title, abstract, author, keywords, unique identifiers, and physical attributes (such as the spatial and temporal extent of coverage). Structural metadata describes the versions and relationships of data to other data (essentially providing a blueprint for how different datasets relate to one another). The score is the sum of the metadata types provided (Table 9).

Table 9. Categories and scores based on the types of metadata provided.

Category	Definition	Score
No or Unknown	No metadata, or data were not publicly accessible so it is unknown if metadata were provided	0
Administrative	Metadata included administrative information.	1
Descriptive	Metadata included descriptive information.	1
Structural	Metadata included structural information.	1

Metadata Standards

Data standards are the rules by which data are described and recorded. The adoption of data standards enables similar data to be have the same format between different agencies, making it much easier to use data provided across platforms. Similarly, metadata standards are the rules by which metadata are described and recorded. Here, we simply noted whether the metadata included the standard used (Table 10).

Table 10. Categories and scores based on standardized metadata.

Category	Definition	Score
No or Unknown	No metadata standards provided, or no metadata provided, or data were not publicly accessible so it is unknown if metadata standards were provided.	0
Some	Some data provided metadata standards.	1
Yes	Metadata standards were provided within the metadata.	2

Timeliness

The timeliness of data provision affects its usability for different purposes. There is a clear distinction between data that are continuously collected on a clearly defined, regular interval (minute to annual) and discrete data that are irregularly collected at an unknown interval. Here, continuously collected data with a defined interval were considered to have greater usability (Table 11).

Table 11. Categories and scores based on how frequently data were updated.

Category	Definition	Score
Unknown	The frequency of updating the data was not specified or self-evident.	0
Not Applicable or Varies	The type of data is relatively static; such as aquifer or watershed boundaries, or refers to one-time data collection efforts. Varies refers to different datasets updated (not collected) at different frequencies.	1
Yearly or Higher	Data were updated less than once a month but at least once a year.	2
Weekly or Monthly and Higher	Data were updated between once a week to once a month.	3
Daily or Higher	Data were updated at least once a day.	4

Length of Record

The length of record indicates how much of the dataset was accessible. This metric is most applicable for time series data; however, it can be useful to have access to older versions of spatial data. Some platforms only provide a current snapshot of the data, often designed to assist users in day-to-day operational decisions. Other platforms allow users to access a portion or the full period of record. Platforms were scored with the assumption that access to the full data provides maximum flexibility for how the data could be put to use (Table 12).

Table 12. Categories and scores based on the length of record that was accessible.

Category	Definition	Score
Unknown	Data were not accessible.	0
Not Applicable or Varies	The type of data was not a time series or the length of record available varied between datasets.	1
Current only	Only current data were accessible. This typically refers websites that provide data on current conditions.	2
Limited Record	A limited period of record was accessible.	3
Period of Record	The full time series for the data was accessible.	4

Reusability Metrics

Reusability was inherent in the nature of this inventory because it is exploring public data and access to those data mean they can be repurposed for secondary uses. The ability to correctly reuse data will be tied to already collected metrics of findability, accessibility, and interoperability.

Total Scores

The total score for findability, accessibility, and interoperability were the sum of the scores of their respective metrics normalized by the maximum possible score. This creates a range of zero (no FAIRness) to 100 (maximum FAIRness). The interactive data visualization tool enables users to compare scores of entities within the selected inventory and across inventories (Figure 4).

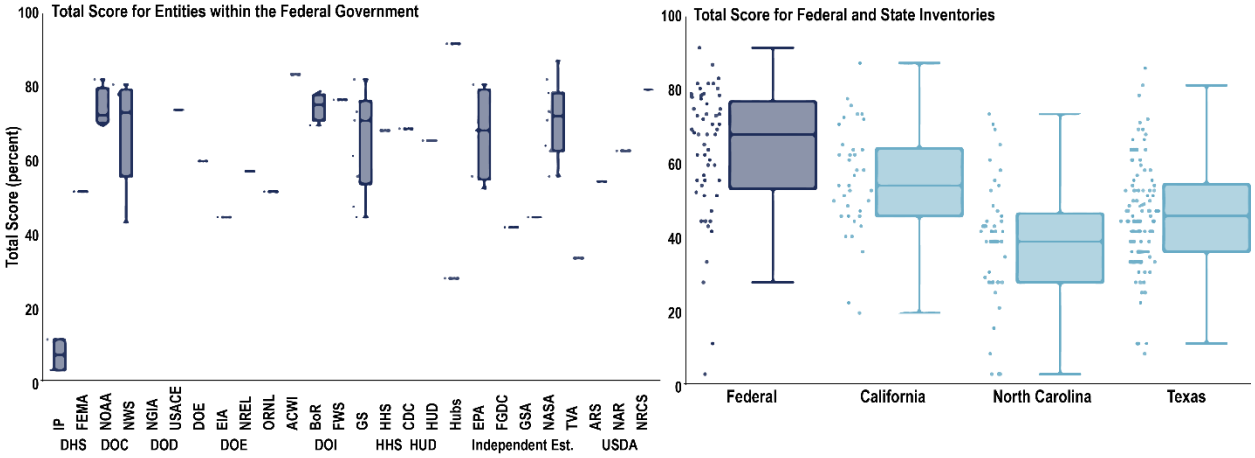

Figure 4. (Left) Total openness score for entities within the federal government. (Right) Total openness score for federal and state inventories.

Results

Users may access the inventories, data, and results here: http://internetofwater.org/public-data-inventory/. The user selects an inventory in the information tab and then explore the results to see who is collecting data, why the data are collected, the types of water data collected, and the FAIRness of the data. Results for the four completed inventories are summarized below.

(1) Who is collecting water data?

The inventory included 268 entities collecting water data and providing those data on 232 platforms (Table 13). The federal government had 42 entities collecting or providing water data. There were nine departments and five independent establishments beneath the executive branch collecting different types of water data to meet their mission. For instance, the Department of Commerce collects precipitation and meteorological data to provide weather and river flood forecasts through the National Weather Service. The Department of Defense collects water data related to the nation's built infrastructure, particularly reservoirs and levees. The Department of Interior collects the majority of water quantity and use data through the BoR and the USGS. EPA collects primarily water quality and utility data. NASA uses remote sensing to estimate changes in groundwater, algal blooms, flood extents, evapotranspiration, and so on.

Table 13. Number of entities collecting water data discovered at multiple platforms.

Government	Overseeing Entity Collecting Water Data	Sub-Entities Collecting Data	Total Number of Entities	Data Platforms
Federal	13	29	42	56
California	2	59	61	34
North Carolina	6	36	42	45
Texas	27	96	123	97
Total	**48**	**220**	**268**	**232**

California had 61 entities collecting water data. California's Environmental Protection Agency, particularly the State Water Control Board, collects the majority of water quality and water rights data. The State Water Control Board has at least 15 programs collecting water data. The Natural Resource Agency collects much of the infrastructure and water quantity data, particularly within their Department of Water Resources.

There were 42 entities providing water data in North Carolina, with the majority of the data provided by the Department of Environmental Quality. North Carolina had a single department oversee the primary collection of all four categories of water data types (quantity, quality, use, and infrastructure). North Carolina consolidated water under a single department in 2014 to improve efficiencies in water management.

Texas had 123 entities collecting water data because their executive branch includes river basin authorities (16 of which provide water data through various programs within the authority). Despite the multitude of entities, the Texas Commission on Environmental Quality (TCEQ) and the Texas Water Development Board (TWDB) collected the majority of water data. TCEQ focused on collecting water quality data for regulatory purposes and TWDB focused on collecting water quantity and use data for planning purposes.

(2) Why are the data collected?

Most entities within the federal or state government collected data for broad scale decision-making (Table 14; Figure 5). Many state entities also collect data for regulatory purposes, while few collected data explicitly for research and innovation purposes. In contrast, less than one-third of federal entities collected data for regulatory purposes, while more than 60% of entities collected data for research and innovation. Between 30 to 46% of entities within federal and state governments collected data for day-to-day operations.

Table 14. The number of agencies collecting water data by purpose. An agency may collect data for multiple purposes.

Government	All Purposes	Day-to-Day Operations	Broad Scale Decision-Making	Regulatory	Research and Innovation
Federal	42	19	35	12	26
California	61	18	49	38	14
North Carolina	42	16	32	29	8
Texas	123	57	101	57	21
Total	**268**	**110**	**217**	**136**	**69**

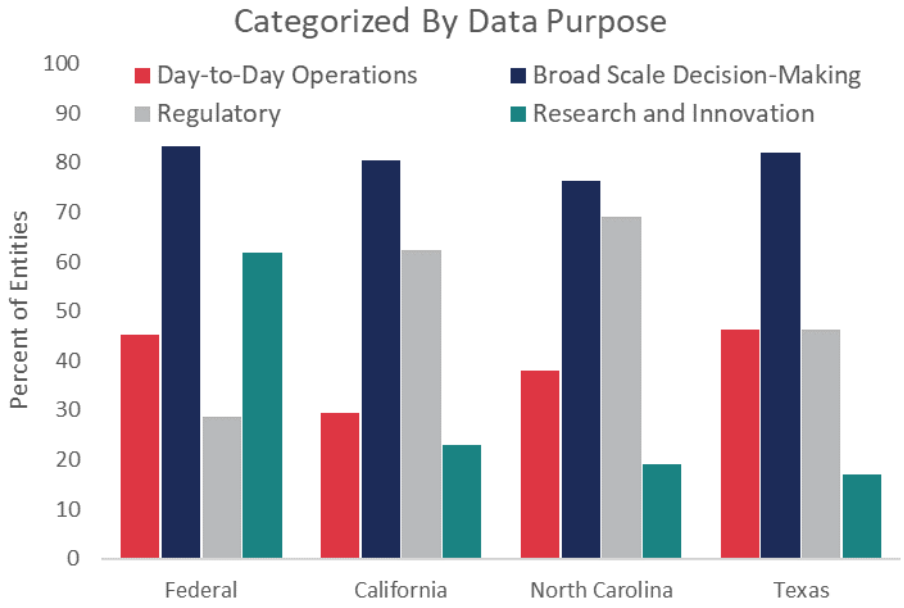

Figure 5. Percent of entities collecting data for different purposes.

(3) What types of water data are collected?

The frequency by which a type of data is provided may indicate the relative importance of those data to a public entity. It may also reflect the method by which data are provided (scattered across multiple websites or located in a single catalog). The majority of federal platforms provided quantity data (76%; Figure 6), particularly precipitation and other meteorological data. Slightly over one-third of federal agencies also provided water quality data (36%) and fewer than 24% provided either water use or infrastructure data. In California and Texas, slightly over 40% of public agencies provided water quantity data, perhaps reflecting water scarcity challenges common in western states. In contrast, the eastern state—North Carolina—only provided water quantity data at 23% of platforms. However, 63% of the platforms in North Carolina provided water quality data (similar to California), only 37% of platforms in the federal government and Texas provided water quality data.

State platforms were more likely to provide water use data (between 34 to 46%) than the federal government (24%), particularly water utility data and data used to support water management plans.

California also provided large amounts of infrastructure data, perhaps reflecting the complicated built and jurisdictional systems responsible for distributing water around the state.

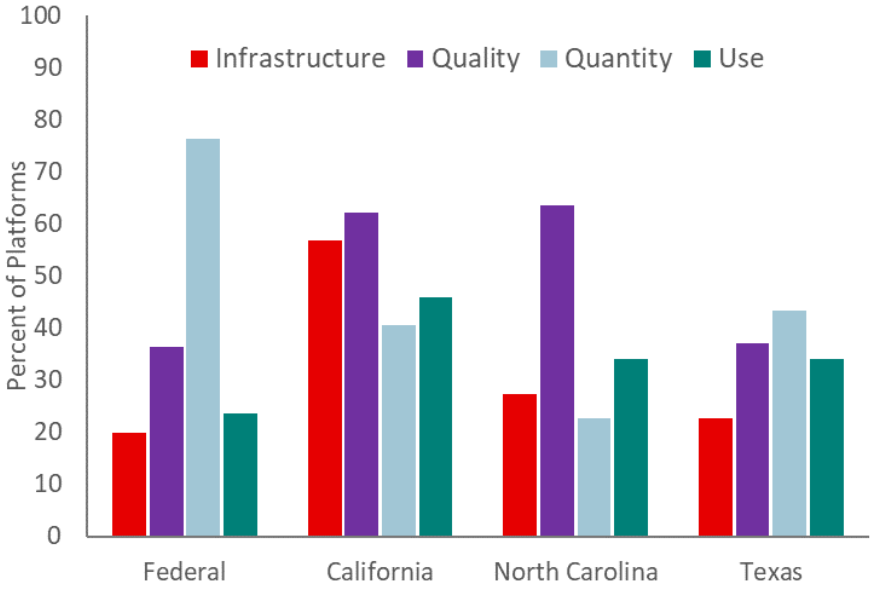

Figure 6. Percent of the data types provided by data access points.

(4) How discoverable, accessible, and usable are those data?

The federal government has made a concerted effort to improve data openness since 2013 and scored higher in findability, accessibility, and interoperability than state governments (Figure 7). The federal government tended to excel in making data discoverable through catalogs and maps, while several state entities provided data on various independent websites. All public agencies performed similarly for data accessibility. Data were often accessible, but the method to access the data varied considerably with some platforms requiring copy and paste or individual downloads, while others have adopted web services.

In terms of interoperability, most states provided data in machine-readable formats with the exception of North Carolina, which tended to rely on proprietary software, such as provided by Microsoft or Esri. Several North Carolina platforms embedded data within pdf reports, and only four platforms provided data in open machine-readable formats (excluding platforms using copy and paste from html). The federal government, followed by California, were most consistent in providing data definitions. The federal government and California were also more likely to provide metadata. Interoperability was the metric that showed the most room to grow for agencies to move towards data openness.

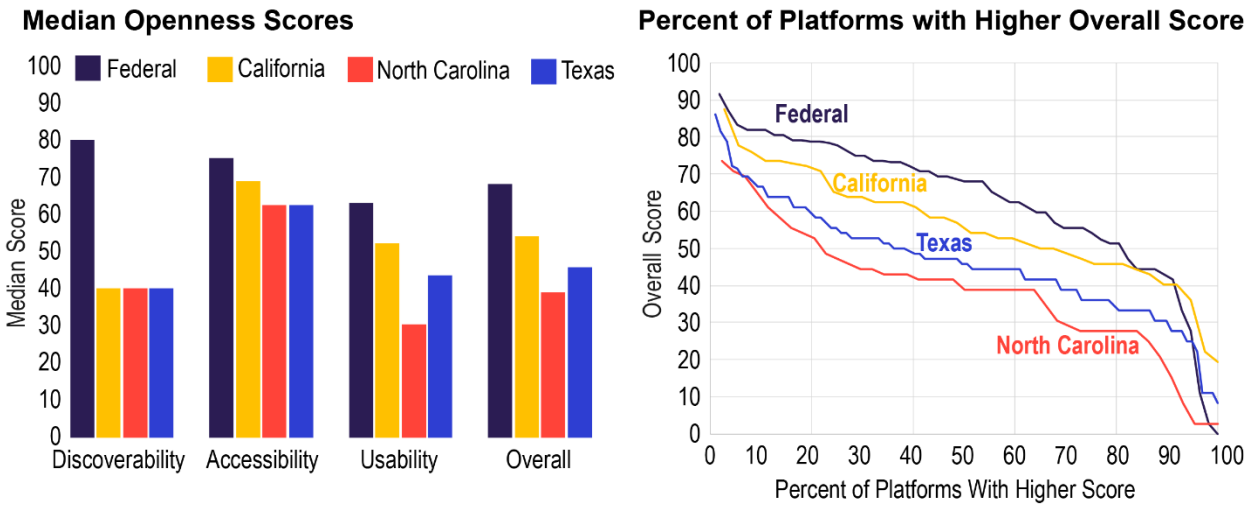

Figure 7. (Left) Median scores for findability, accessibility, interoperability, and overall for each data inventory. (Right) The percentage of platforms in each inventory with a higher overall score. For instance, 80% of federal platforms had a score greater than 50, while only 22% of North Carolina platforms had a score greater than 50.

The majority of federal platforms had an overall openness score greater than 80 (52% of platforms) (Figure 7) with 14% scoring higher than 80. Nearly 43% of California's platforms scored higher than 60, while only 20% of Texas and 14% of North Carolina platforms had a score greater than 60. This demonstrates that while the median scores are somewhat comparable between inventories, in the majority of federal and California platforms had relatively high openness scores while most of North Carolina and Texas platforms scored between 30 and 60.

There were 39 platforms with an overall openness score higher than 70 (24 in the federal government). Two of the highest scoring platforms were inter-agency efforts between federal agencies (Table 15). The Water Quality Portal scored at 91.7 and the National Groundwater Monitoring Network scored 83.3. These portals allow federal, state, and other entities to share their data through these portals, provided the data are standardized.

California's Natural Resources Agency Open Data platform had the highest score of any state platform (87.5) and the second highest score overall. Similarly, the Texas Water Development Boards (TWDB) had two open data platforms that scored higher than 80: TexMesonet platform (86.1) and the Water Data for Texas platform (81.5).

Table 15. Platforms scoring higher than 80 in openness for the federal government and higher than 70 for state agencies. Many federal government platforms scored between 70 and 80. See the online inventory for more details.

Inventory	Organization and Platform	Total Score
Federal	Joint: Water Quality Portal	91.67
	NASA-ESD: National Snow and Ice Data Center DAAC	87.03
	DOI-ACWI: National Groundwater Monitoring Network	83.33
	DOI-GS: National Water Information System	81.94
	NASA-ESD: Socioeconomic Data and Applications Data Center	81.94
	DOC-NOAA: Climate Data Online	81.94
	EPA: EnviroAtlas	80.56
California	NRA: Natural Resources Agency Open Data	87.50
	NRA-CGS: DOC Maps: California Geological Survey	77.78
	GOA: CA Open Data Portal	75.92
	EPA-Integrated Water Quality System Project: Integrated Water Quality System Project	73.61
	NRA-DWR: Water Data Library	73.61
	EPA-GAMA: GAMA OnLine Tools	72.69
	NRA-DWR: Water Management Planning Tool	72.22
	NRA-Ecosystems Conservation: Biogeographic Information and Observation System	70.83
North Carolina	IT-GIS Data: NC Onemap	73.61
	DEQ-Online GIS: NCDEQ Online GIS	70.83
Texas	TWDB-TexMesonet: TexMesonet	86.11
	TWDB-Water Data for Texas: Water Data for Texas	81.47
	TWDB-Groundwater Monitoring: Groundwater Database	78.69
	TRA-Planning and Environmental Management: Lake Livingston Contrail System	72.22
	TDIR-Statewide Data Coordination: Texas Open Data Portal	71.31

Final Thoughts

These data inventories provide a broad overview of the types of water data collected by federal and state entities and the openness of those data. This inventory is a living document and can include more states, and/or be updated periodically over time. The inventory begins to create a framework to demonstrate the spectrum of findability, accessibility, and interoperability and how to progress to greater levels of FAIRness. The IoW is developing resources (http://internetofwater.org/resources/) to assist state agencies to improve the openness of their water data.