

Advisory Board Meeting

February 7, 2019 Conference Call

Board Members Present (7 of 9)

Jerad Bales, Kelly Bennett, Martin Doyle, Peter Colohan (ex officio), Sam Hermitte, Sara Larsen, Emily Read, Dwane Young

Updates from Chair and Executive Director

The team has spent the past several months securing efficient funds to make hires and launch the IoW. We made our first hire of Peter Colohan as the Executive Director for the Internet of Water (IoW) starting on December 1st. Martin Doyle will be the Advisory Board Chair while Peter Colohan is charged with implementing the strategic vision and implementation plan that we will be discussing today.

Colorado River Basin Roundtable Outcomes

The Aspen and Nicholas Institutes hosted their last IoW roundtable event for the Colorado River Basin January 29-30. The roundtable discussed three ongoing water data initiatives that have relevance to the Colorado River Basin: IoW, Water Data Exchange (WaDE), and the OpenET project. The roundtable provided clarity on the needs of the community for the IoW to serve as a head librarian that provides a catalog of water data and serves as a convener to discuss standards, best practices, tools, and so on for various stakeholders and communities. The major felt needs were to (1) build a catalog/clearing house of data, standards, and definitions, (2) provide coordination and convening of efforts and resources, (3) create best practices for cataloging, metadata, standards, and so on, and (4) provide technical resources. The desire of the community was not for the IoW to resolve issues, but to provide a platform to unearth the variability and have a conversation without value judgments.

There was consensus that a dictionary of terms and what they mean to different stakeholders would be of value, without spending time and effort on trying to normalize the data and terms. There would be immediate value to many communities to begin building a catalog of data and definitions/terms as a reference resource for everyone. A specific data type in need of cataloging definitions is water use data, something that WaDE is working on with USGS. This would also be relevant for the Colorado River Basin as it seeks to clearly define and document consumptive use prior to renegotiation of the Colorado River Basin Compact. This is an aspirational effort and a big lift, but we do want to start building a library that anyone can access.

Discussion of IoW Strategic Plan and Key Concepts

The original Internet of Water report from Aspen forms the basis of our strategic vision and we seek to stay within the bounds of the consensus of that diverse group of stakeholders. During the Board's

review of the Strategic Vision there were a few key areas of diverging opinions that we want to discuss as a group.

Open and/or FAIR Data

What terms should be used to describe the IoW: open and/or FAIR (findable, accessible, interoperable, and reusable). The proposal was to use the term “open and FAIR” and include those definitions as footnotes. The definitions have overlap but are slightly different with open data being broader and more encompassing than FAIR data. Not all the data in the IoW will be open necessarily – we can include non-open data if we are creating a catalog, but I think we want to preserve the goal and notion of openness. We can link to IoW Principle 2.4 that provides recognition of the importance of ensuring data privacy and security. What are your thoughts?

The consensus was that promoting water data to be open and FAIR is fine coupled with Principle 2.4. In addition, open can be used as a noun, but not a verb. We promote open data, but we need to use a different term from opening data. Perhaps, we can expose data or make data findable? This also avoids the connotation that data are somehow locked down. There was also acknowledgement that there needs to be good communication and education to general audiences about what terms like “open” and “FAIR” mean.

There was a proposal to consider adding to the FAIR data a geofabric component for water – that water data will be georeferenced and attachable to a watershed, aquifer, etc. There is a federal effort to link water data and make it spatially searchable. Perhaps these could be Water FAIR or FAIR Water principles? Perhaps this can be a progressive evolution as we describe principles and show examples of best practices and standards in the water space that align with those principles.

The board agreed we should stick to the basics and focus on our fundamentals. The internet is at its core, a library catalogue and index, and that is where we should start. We need to identify what data exists, in what format, and provide links to the data. That would be incredibly impactful as my organization receives frequent phone calls asking where to find different data. Is open and FAIR where the IoW is at right now, or are we getting ahead of where we are at right now?

An agreement that the fundamental thing we must do is create water data catalogues. There was discussion around whether (1) the IoW should / could build such a catalog or (1) if hubs should build the catalog and the IoW provides a framework to build towards catalog uniformity? A concern of asking hubs to do the cataloging, or even asking them to adopt principles, is that it places a burden on them and inequity in the ability of hubs to meet that call. Large hubs would have the resources, but smaller entities would not. Another option to cataloging may be to use search engines once hubs provide data in a standardized format that is machine readable and following standardized rules of access. The IoW has a tremendous opportunity with a limited window, time, and funding to create a consolidated place to see what data exist. What is more necessary is for us to be able to create a place to just know what the data are, where they exist, create linkages to access those data in as automated a way as possible. The IoW should think about how to expedite and remove the main barriers to the willingness to participate.

While standards and principles are important but there is a practical challenge to really starting there. This will be a focus of conversation at the next IoW Board meeting.

Lead Partner / Anchor Tenant / Coalition Concepts

The board also explored how to ensure the long-term viability of the IoW. One idea suggested was to build large national coalitions, however it seems that might be too ambitious and not desired by the community. There may be a possibility/need for some form of coalition at the local scale to ensure the success of hubs. Can the IoW be more than a product sponsored by public dollars for the public good? Is there a business model that can encourage the private sector companies that could be anchor tenants benefiting from IoW efforts to help fund the IoW? Coalition building was one idea, but it seems that there might not be a demand signal. Who are the private sector anchor tenants? One suggestion was a coalition of water utilities or the U.S. Water Alliance. Other ideas include the energy sector and corporations with water intensive infrastructure, such as data centers. Private companies interested in ensuring they can find the data needed to assess their water security are likely anchor tenants. Another consideration was to explore the value of regulatory relief to the private sector.

The general consensus was that there is not a demand signal for the IoW from the private sector (data user) perspective. The savvy private sector water data folks would benefit from the IoW because it makes it easier for them to find and access water data, but they probably wouldn't pay for that service. Corporate philanthropy seems to be the most viable option, but that is not a sustainable funding stream. Perhaps a consortium of entities donating to the IoW for the public good could become sustainable.

Another option for private dollars to enter would be for the IoW to provide value added products. For instance, if the IoW built a great catalog of library data where folks can get the raw data freely, but then began curating some datasets to create aggregated, clean datasets that individuals could purchase. That could be one avenue to get private dollars participating in supporting the IoW; however, depending on the dataset, that might only cover the costs of the added service of data curation. Along those lines, if the IoW built the basic infrastructure and incentivized standardization of water data, it is possible that the private sector would support the IoW because they it better enables them to create data products at scale.

Perhaps the IoW should take time to assess other business models. For instance, the World Resources Institute (WRI) created the Aqueduct Map as a derived product to assess water risk. Bloomberg financially supports WRI and ingests their Aqueduct Tool into the Bloomberg system where they have thousands of subscribers paying \$24,000 annually to access the combination of water and financial data. Bloomberg is probably unique in that they have locked down certain datasets and have pursued an aggressive funding model that might not be possible today. Another example would be to explore the Open Geospatial Consortium (OGC) and understand their financial model and how they built a community of individuals interested in standardizing different data types and monetized it. There are likely a handful of large organizations that might support the IoW, but the funding model approach we take will be very important.

This is an open conversation that we will continue to return to and hopefully gain clarity as the IoW begins producing.

IoW Hub Concept

It is important for the IoW to clearly articulate what constitutes a hub. What is an IoW hub and what does it do? There was a lot of agreement around this term being used in the broad community to mean different things. It should not be the role of the IoW to define who is or is not a hub. That becomes prescriptive. However, it is important for the IoW to define what an IoW hub represents. The Aspen Report defined a data hub to be: “a formalized, structured source of federated open water data.” The board discussed other possible definitions:

Definition 1: A hub is a place where a diverse group of stakeholders contribute data. The hub is part of an organization. For example, CUAHSI organizes academic data and WaDE organizes state water use data. It is unclear that federal websites that are only providing their own data would constitute a hub with this definition.

Definition 2: A hub provides a single point of entry to discover and access data provided by a multitude of sources, as well as some level of value added products. This could be improved search capabilities or visualizations of the data. The challenge is that this definition excludes federal agencies that only provide their own data, but provide vast quantities of data across the nation, often using a multitude of sensor types.

Definition 3: A hub federates data and builds a catalog. The building a catalog could be part of what sets the vast world of hubs aside to become an IoW hub.

Definition 4: A hub has to champion a particular type of data or a particular group of stakeholders. It could be similar to the Open Water Data Initiative (OWDI) model where the OWDI would sign a MoU with a group that could demonstrate a successful working pilot. The IoW could define IoW hubs through MoUs with a hub that successfully demonstrates some set of criteria.

Definition 5: A hub is a single access point for a type, or some types, of data and those data are standardized and made accessible to users. Are hubs the underlying software and infrastructure or are they an organization with leadership? A hub brings data together and that typically requires many partners such that no single organization should claim credit for a hub. For example, the USGS provides streamflow, water quality and groundwater data. While 85-90% of the data is from the USGS, the other 10-15% comes from other agencies that have gone through our quality assurance process. There does need to be leadership for a hub to be successful, but the foundation of a hub is the system itself.

Definition 6. The state or regional level is the right scale for a hub that could demonstrate broad, societal value. Hubs should not encompass a smaller regional area. From a state perspective, Definition 2 is ideal as states hold a large variety of data. Our critical need is for there to be a catalog of the data within a State so we can simply find the data and to make those data relevant and accessible to the

general public (not tech savvy users). A hub connects diverse data that would otherwise not be connected. I think a node would be a single data source or type.

Definition 7: There seems to be a need to introduce new terms to understand if a hub is the people and geography or data types and standardization. Perhaps a hub is a location that retrieves collections of data of a single type: sensors, samples, administrative, and so on, from a variety of different producers. Perhaps an IoW hub provides some level of standardization of the data and APIs so data can flow seamlessly between hubs. For instance an eastern state sensor hub would be able to talk with a federal or western state sensor hub. Hubs should not be catalogs. A catalog is a registry of available hubs or datasets and how to find those data. Lots of states have catalogs already – Colorado is a great example (as is the city of Austin, TX) where all the datasets are indexed and discoverable. There is a need to standardize catalogs between agencies. Perhaps a coalition can be a term to describe the people that collaborate around a hub or a catalog. A critical component of an IoW hub should be that the data are georeferenced and have hydrologic search functionality.

There was pushback on this definition as the Aspen report clearly says to steer away from prescribing standards as it will exclude so much data and so many people from participating. There was agreement the IOW cannot be prescriptive and participation is purely voluntary. At the same time, we want to accept all data but move towards interoperable standards to become IoW compliant. In essence, the IoW would provide a roadmap to becoming an interoperable IoW community.

Definition 8 Perhaps Definition 7 is primarily referring to a node, or a single type of data, whereas a hub is an organization that is governed by a group of people that are creating a platform or mechanism for federating data. Node seems to be technical while hub includes people. Perhaps a coalition describes folks managing hub. Nodes are the subsets of data that the hub is federating. For example, the Clean Water Hub is actually a node within the citizen science community hub of all types of monitoring data.

There was pushback here. A node is a publisher of a type of data by a single entity (data producer). A hub is a collection of nodes. WaDE is a hub and the states are a node publishing water rights and appropriation data. The Water Quality Portal is a hub that receives data from multiple nodes using the WQX standard. If a hub is not a clearly defined collection of data then anyone could create a website with a bunch of csv files for sampling data, sensor data, and so on. What is to prevent that entity from being considered a hub?

Under this definition, a state would need a hub for water use data and a hub for water quality data with a larger umbrella over these distinct hubs. It seems that under definition 8 you couldn't have state hubs. Having a hub focused on a single data type may make sense at a federal and regional level; however, that definition won't work for states. Siloes of data by type are part of the problem states are wrestling with. One agency is collecting water quality data while another is working on water quantity and another on water rights. There is no fabric connecting these types of data together and that is a desire for the State. Colorado provides more of a catalog than a hub, but that is what a state needs first and foremost to know what data even exist at the flood control district, local utility, irrigation district levels. Many states do

not know what is out there. We have aggregated data from these communities but no easy way to get at the underlying raw data. From a state perspective, we need the catalog first and maybe the coalition is the fabric pulling the catalog together. Perhaps the next step would be a hub.

Many states are using Esri's open data portal (which they call a hub) to share data. Perhaps a hub can be at the state level and it is an assemblage of data that agencies want to share and the specific agencies sharing the data are the node to the state hub?

Consensus was to spend a few weeks coming up conceptual models and frameworks for hubs to bring to the in-person advisory board meeting where we will hopefully come to some resolutions.

Next Steps

We will set up monthly calls to continue figuring out programmatic terms and operational frameworks that will be the foundation for how the IoW builds over the next few months.