AB 1755 Online Public Meeting and Reception:
Launch of California Water Data Consortium
June 26, 2020

Notes from Breakout Groups

During the second half of the meeting, participants separated into eight facilitated breakout groups to dialogue around topics related to the Consortium and water data. Facilitators led a round of introductions and then sought to identify if there were any attendees in their group who were new to the statewide water data effort, so they could choose questions geared to their participants. They then engaged the group in discussion on two of the following three questions. Below are compiled responses.

Question 1. Is data important in your work?
[For those less familiar with the Consortium]

Breakout Group #2
- I am with The Nature Conservancy and have an affinity for data, and deal with data around groundwater and surface water flows, diversions, streams, and groundwater.
- I am a technical lead at DWR for the National Hydrography Dataset Stewardship Program, ensuring the California portion of the NHD is the best it can be. I would like to help link this State data with NHD; and GIS data.
- I am at MNWD, which has an awesome GIS team and data science group.
- I am a statistician, I don't use water data directly in my profession, but I work with folks at NOAA CREST (Cooperative Remote Sensing Science and Technology Center) using USGS streamflow data and climate data to predict surface water levels/availability.

Breakout Group #6
- Snow data expert. I have done a lot of work with data cleaning, trying to come up with generalized methods for data cleaning the 1.36M observations.
- Making data useful for making informed decisions. Partnered with MNWD on stormwater management. There is an over-abundance of runoff when it's raining, so if we could do something with it…. Urban Drool is also an issue. How might we utilize stormwater in both wet and dry times?
- Water District. I use State data (snow, precipitation, and evaporation) and want to plan for water supply. State datasets already exist, so the question is how this effort helps. Same question based on reporting. Can we meet multiple reporting requirements with a single measurement effort?
- Non-profit involved in river restoration. I deal with the intersection of water and agriculture and have been working with wrangling datasets from all over and

reformatting. We do lots of modeling and it's a big effort each time. Developing a visualization tool for pesticide impacts on surface water health (ecotoxicology and human impacts). I understand the Consortium isn't looking to build tools initially, but would like to see this after the database component is sorted out.

- Water Boards. I collect surface water data in my region, SWAMP (Surface Water Ambient Monitoring Program) program. I want to find ways to have better QA for the data going in and a better way to utilize the data for visualization by both staff and the public.
- Academia. I heard about the Consortium through the water data challenge. I enjoyed seeing the community around water data, an aspirational group.

Question 2. What would you like the Consortium to achieve in the next year?

Breakout Group #1
- Define authoritative data sets, and clearly pick which agency is a champion for that dataset - even that subagency. For people who want to find the authoritative dataset – offer a one-stop shop.
- Integration of habitat components included in online databases, like USGS – it collects surface water data, and more information about groundwater. A lot of fish information integrated into easily obtainable databases – salmonids, other listed species.
- The process now: find data on a website; pull it a single time, using clicking around. It would be great if there was an API to put into code – I know that's down the line. It would be cool to see one pilot project taken to the nth degree. REST (representational state transfer) API is able to handle a lot more traffic and be hit and updated repeatedly, unlike a database. There are different silos talking – a better use case is a machine-readable interface. A better case is blurring the lines.
- A first version of the quality standards – I want a pragmatic version, not the gorgeous final version, but basic agreement on basic metadata, basic units, basic names to define groundwater, how we're going to flag the calibration data, agreement on core representations, so the early data as you pull it in can have a prayer of being integrated at all. Typically, if you changed the way you gathered it, and standardized it, it would have been 100% easier to integrate.
- The GAMA OnLine Tool site (Water Board's Groundwater Ambient Monitoring and Assessment Program, https://www.waterboards.ca.gov/water_issues/programs/gama/online_tools.html), is a best practice. That is exactly what you guys are talking about – 7 to 8 screens you can tap on and integrate everything. Programmers are putting it into a universal system; they can talk to each other - oil, groundwater, etc.

Breakout Group #2

- Streamlining is a top priority!
- The relationship between groundwater, aquifers, and water supply.
- There is a lot of room for improvement, which is exciting. The main thing would be to avoid the "build it and they will come" mentality. Take a user perspective. Who is using the data; who should be? Focus groups would be informative. A lot of data is open – skills are often the limiting factor.
- Streamgage data – USGS has a good model for compiling data and how to back out to API. DWR's streamgage data is good, but could follow the USGS model.
- I work on water data challenges. There is a need to support folks (students, etc.) in getting started in challenges and other community engagement. Data can be overwhelming, and the relevance of various datasets can be unclear. We might benefit from a "Get Started" guide. What would actually help a given State agency? What are the biggest funding issues, biggest problems, etc.? Where to look, how to access. Instructive videos on a YouTube channel would be best!
- Streamlining – clear reporting and <u>metadata</u> requirements by data type. The Consortium is best poised to gather input from stakeholders.
- More focus on user need definition, focus groups. We've done a lot, but there's room for improvement. Go alone to go fast, go together to go far. Better preparatory packages for data challenges; use cases to document what we'd like to know.

Breakout Group #3

- Streamlining of data. It may require some statutory changes; how can the Consortium move in that direction, identify?
- More protocols of data sharing (machine readable) and prototypes to tee up the agenda for the larger community to get interchange data – real examples, try, learn, iterate.
- Building relationships among philanthropy, academia, private, public, community.
- How to codify standards, more development.
- Streamlining in other data areas beyond groundwater (such as water quality).
- Metadata protocols and development.
- Acknowledging how existing data structures and sources of data fit into the strategy.

Breakout Group #4

- Build out use cases - how to collaborate with existing groups.
- Plan for integrating federal partners. I would like to see how to leverage data so it doesn't lead to duplication or double-counting.
- Another idea – great as part of a publications strategy – use a digital objective identifier; I would love to use it.
- For the Legislature: Easily digestible briefs; that's really helpful to reinforce and advocate for transparency.

- Code for America - I am on the brigade side. I am always looking for new areas to work on – I don't know enough about water, but I'd like to get involved.
- Make sure you do broad outreach, and include groups who aren't already part of the Consortium, such as the Groundwater Resources Association, cities and counties, etc. Do broad and frequent outreach; spend time on the mission of the group, not just the group itself.
- Interested in what we can accomplish with the data – tied to the use cases – we'll never capture all that we need, but to the extent you can come up with more or varied ones – let a high-schooler in a disadvantaged community (DAC) do one, not just for the State folks.
- Continuance of collection of data; as we go into a round of budget cuts for our programs, data is not always appreciated - it's important to advocate for the value of data.
- Totally agree – and how to deal with uncertainty.
- At all times, but especially now, we should demonstrate the value of water data to disadvantaged communities. So I encourage that we bear a use case in mind that foregrounds DACs.

Breakout Group #5
- Initial projects should focus on aligning efforts across agencies and programs, streamlining data submissions (e.g. surface & groundwater, diverse types of data).
  - The role of **standards**: the Consortium could help in discovering any inconsistencies with standards; it's an opportunity to *improve* data standards.
  - Emphasis on representing users, State and Federal groups.
  - Even after standards and a theory for streamlined processes exist, getting to practical implementation can be difficult -- resources, best practice guides are needed to get the implementation to happen.
- We would like to see documented examples of current duplication/redundant reporting, and use a feedback loop, multi-stakeholder iterative process.
  - Notes initial work on consolidation.
- Incremental change; if there were a standard, we would be better positioned for new incoming data. We are willing to make the effort to implement long term change!
  - Notes best practices arise from multiple examples, concrete use cases.
- Notes issues in translating across disparate datasets (e.g. chemical and trade name of pesticides); improvement in data dictionary consistency is critical.
  - Agrees.
  - Agrees also on challenges with translating across datasets, e.g. lat/long framework.

Breakout Group #6

- Describes the evolution of silo'd reporting – it makes reporting onerous and has a lack of transparency on usefulness. "How can we take all of these siloed datasets and standardize them for use beyond reporting?" In the coming year, we need to make a focused effort on communicating that the ability to visualize data is a priority of the Consortium's effort – we should lay the roadmap out clearly for those who haven't followed as closely.
- There are lots of data and datasets, but I have no idea what's out there.
- I would like to see this too.
- https://github.com/ropensci/weathercan is an example of the ability to query from R that tells you what's out there, the data.

Breakout Group #7
- 1. Streamlining to data submission. This has been attempted without success by a variety of groups both internal and external to the State without much success, but the Consortium seems uniquely positioned to tackle this issue.
- 2. Facilitation of the development of standardized protocols, formats, and methods of providing data based on topic area.
- 3. Identification of development of topics of frequent interest that can be turned into decision support tools by either the State or outside entities.
- 4. Recommendations of improvements to the existing State portals and how those portals can be changed to better serve the needs of data users.
- Two people identified themselves as being interested in data streamlining.

Breakout Group #8
- The Consortium provides flexibility to facilitate existing collaborations that exist to take the data that's already being created to increase value at a watershed scale.
- We've created many data tools at various demonstration projects funded by Bechtel and others. These tools could be extended to other places and projects, should be elevated by the Consortium.
- Gap analysis on the kinds of data and monitoring, especially on the watershed scale. It should start with basins where there's a pathway to succeed. What's required to bring these basins into balance? There's a tendency to cut science first in tough budget years. Good basins to start with: North Coast rivers, Mokelumne, Cosumnes.
- It's all about results. It's important to have a place-based approach, but it also needs to be decision-based. It needs to be scalable. I am currently working with the State Water Board on water data. Efficiency is great, place-based decisions are great, but it might not change the outcome. We need to think about decisions that need to be improved that are inherently scalable.

Question 3. How will we know when we've succeeded?

Breakout Group #1
- A timeline by this point in time, a detailed timeline/schedule for various data. A roadmap, and write in their requests.
- To show that there is an authoritative – a place in the portal where in a few clicks I could get to the authoritative data library. One point of contact to open all the other doors I want to get into.
- The GAMA site has that info – it is very up-to-date. I found data entered in May of this year.

Breakout Group #2
- Key performance indicators (KPI) can be difficult to determine, but well worth the time.
- State agencies aren't always well resourced. We count on partnership with other agencies and stakeholders to get done what needs to be done.

Breakout Group #3
- The number of citations of the data published (reused, used) – beyond the loading of the platform "hits," including citations in the reporting of the used data to the government (compliance and other things).
- A prototype of a machine to a machine compliance data flow to demonstrate potential, find issues, learn from it, etc. If you accomplish this, you would have end-to-end visibility of this automated compliance.
- Life cycle of data flow to better understand the burdens and pain points to "data collection" and then use root cause analysis to identify potential actions to improve (or reduce burdens).
- In the first year, we should try to identify the redundancies and eliminate the ones we can find/attack.

Breakout Group #4
- Look at how many people have heard about the work of the data Consortium. Do broad outreach – it's easy to measure if more are involved after the first year.
- Identify key performance indicators early.
- How to measure collaboration; how to measure success; many different things to measure; are we able to address questions we haven't addressed before?

Breakout Group #5
- Measuring use of (e.g. download of) datasets.
- One measure of success could be legislative acknowledgement, e.g. contributing to an "AB1755 v2".

- Beyond downloading, how can we tell data are useful, how they were used, additional data needed, meeting needs?
- Perhaps there's a way to get feedback on what columns of data within a dataset are requested.
- Notes Google Analytics.

Breakout Group #6
- One criterion for success: rather than managing the data myself, the data steward should emphasize QA so that I can focus on research informed by the best data.