SWADE Agency Workshop Report

On August 3, 2021, the UCI SWADE Team held a virtual online workshop to solicit feedback on the creation of the SWADE water data sharing and analysis platform. The workshop consisted of three hours of presentations (including by the CA Data Consortium and CA Data Collaborative), question and answer sessions and small break-out group sessions to better understand the needs and concerns of water agencies that will comprise the userbase of the platform. The workshop was attended by representatives of the following agencies and organizations:

- Inland Empire Utilities Agency
- Eastern Municipal Water District
- Coachella Valley Water District
- Orange County Public Works
- Orange County Water District
- Moulton Niguel Water District
- Orange County Sanitation District
- Santa Ana Watershed Project Authority
- Metropolitan Water District
- California Data Consortium
- California Data Collaborative
- Los Angeles County Sanitation District
- City of Lakewood
- Los Angeles Department of Water and Power
- California Data Consortium
- California Data Collaborative

What tools would be most useful to you in a data analysis platform?

Many respondents indicated that a key platform component would be an intuitive dashboard where it is easy to upload and access data, there is version control and metadata and usage is well-documented. This should streamline data collection, processing, visualization, analysis and reporting processes and contain algorithms that allow for autonomous operations. This could be useful for: illicit discharges, identifying leaks before a failure, historical data analysis to predict problems. Version control and documented metadata would also be useful (i.e., multiple versions of the same datasets and ensuring the most current version is updated in the system).

Respondents also indicated that tools to help in decision making (e.g., water quality, management practices to deal with pollution sources, assessing whether processes and investments are working) would be useful. GIS functionality for environmental data and infrastructure data is also valuable as well as tools that enable access to other agencies' data. It should be noted that most agencies do not employ data scientists. So, the platform must be easy to operate for a non-data scientist.

What are your concerns or apprehensions with using a data sharing platform?

A concern that was voiced throughout the workshop is that agencies are worried about their data being misused or misinterpreted by outside organizations including the public, regulators and other agencies. Agencies rely on public trust and spend time and resources to message to the public that their water is safe to drink. Any data that could be misinterpreted by the public or regulators to undermine this trust is a primary concern for agencies. Respondents indicated that NDAs with other agencies could work well to address this concern. Additionally, some datasets cannot be shared legally or person to person within an agency, so making it clear what can/cannot be shared within the platform will be crucial.

QA/QC is also a concern as data must be checked and approved by agency employees before being shared. This could impact the real-time functionality of a platform, as data would likely not be able to be supplied in real-time. Privacy and security issues are also a concern for some agencies. However, others indicate that it is less so as a platform/agency would aggregate data to protect individual's privacy. In such as system, customers would need to opt-in and if this is not widely done, customer use data could be less abundant. Some data, however, must be made available to the public in order to comply with certain regulations. Sharing critical asset location data is a security concern as it makes infrastructure vulnerable to hostile actors if accessible.

How can we work together to address your concerns?

The Open and Transparent Water Data Act (AB 1755, Dodd) requires the Department of Water Resources, in consultation with the California Water Quality Monitoring Council, the State Water Resources Control Board, and the California Department of Fish and Wildlife, to create, operate, and maintain a statewide integrated water data platform; and to develop protocols for data sharing, documentation, quality control, public access, and promotion of open-source platforms and decision support tools related to water data. There is also the California Public Records Act (Statutes of 1968, Chapter 1473; currently codified as Chapter 3.5 of Division 7 of Title 1 of the California Government Code) requiring inspection or disclosure of governmental records to the public upon request, unless exempted by law. Therefore, for most of the regulated parties, social concerns are only secondary to regulatory compliance. However, trust that the data won't be misused/misinterpreted is critical. Also, agencies must feel that it is worth the time to get the data into the platform (i.e., the platform must present a significant benefit to agencies

in order for them to share data within the platform that is beyond what must be shared with regulators using existing sharing protocols). Respondents reiterated that there is hesitation to share data in cases where it may reflect poorly on their agency (i.e., exposing what they are doing or not doing). However, participants indicated that regular operational data sharing among agencies may be useful during an emergency.

The water industry is historically risk averse and using new technologies such as water data sharing platforms may be met with resistance. To shift this industry mindset will be difficult. Social pressure on the industry can help toward developing the data sharing mindset and facilitating the next generation of water data leaders to help to grow the data sharing/ exchange mindset can change the current culture. There are many organizations working on water data projects and SWADE should coordinate with others working on similar goals. Participants indicated that there may be an opportunity for the SWADE team to identify data gaps to highlight the problem of data management and utilization and what is at risk. For example, there are 3,000 water providers in CA, yet there is no SCADA for extreme events. Water agencies utilizing advanced data analysis are a minority. However, estimates of risk of economic loss and social equity problems (especially where quantitative), for example, would bring greater transparency and progress to the water industry. One limitation to this type of utilization is that some of the information required to make these assessments is not currently documented in all agencies.

What questions do you have?

Agencies hope to better understand how the repository data gathered by NSF can be used for the benefit of the agency/communities/decision making. To better elucidate this can efforts from the past/repositories be analyzed and applied to our participants. Participants indicated that better understanding what challenges other agencies have would be useful and could help participants work together to improve as a group. Respondents indicated that there are a number of water data sharing efforts occurring at different issues and want to know how we could deepen our collaboration with other entities. They recognize that building the data platform and tools is important, but the SWADE team may also be able to make an impact in other areas of water data science and management. Participants would like to know more about the collaboration anticipated with the CA Water Data Consortium and the CA Data Collaborative (CaDC). They want to know more about how the data will be uploaded into the platform and suggest working with existing organizations such as the CaDC who have tools (such as the OWRS survey at CaDC) for doing this that agencies may already be familiar with.

If you are currently sharing any data with other water agencies, please describe.

Los Angeles County Agencies

LA DWP is sharing data with research organizations and universities mostly for research purposes (asset management). They are also sharing data on water loss efforts/water resources. Within LA city, data are being shared between potable water and wastewater agencies. For example, MWD shares data with LADWP and with all member agencies, including agencies in Northern CA and regulatory agencies.

The City of Lakewood, a smaller entity, is sharing compliance data, water supply, production, quality data. They also receive requests from university professors for specific projects.

Data sharing is done using one drive, private cloud, thumb drives, NDAs or excel. SCADA data, which is less available, is shared through a chain-of-command and is difficult to share even within an organization. Email is often used to send data files. Because staff at agencies changes somewhat frequently, it may be difficult to develop relationships with the appropriate staff at each agency responsible for sharing data, thus making sharing more difficult. Some agencies have restrictions on sharing methods, such as LADWP which does not allow data sharing on any cloud or email service without approval. Sharing energy (power) data has federal regulations. Some, but not all of these regulations, also apply to water data.

Data sharing is not widely used at these agencies simply due to the fact that it is not a high priority (there are not clear benefits). More resources are needed to allow for regular data updating and sharing, but the cost v. benefit is unclear. Agencies also want to make sure that the data they are sharing is consistent and correct, which takes additional time and resources. Additionally, many agencies do not see the political value in data sharing.

Agencies indicated that size of organization often dictates data usage and needs and this should be considered when designing a platform. However, all agencies identified a number of areas/projects where additional data from other agencies would be useful to them. These include: groundwater pumping rates, additional contaminant analysis, reclaimed water/source water data. They suggested that pilot projects to test technologies for water management and tools within a community could be useful as the platform is created.

Orange County Agencies

Many Orange County water agencies are working together on collaborative research and infrastructure projects. They share data to support these projects on an as-needed basis. For example, MNWD had a data sharing project with OCWD to develop a PFAS tracking tool as

part of the CA Water Data Challenge; SAWPA and OCPW are working together to develop a data dashboard and trying to create a data sharing dashboard to reduce paper; and MNWD and OCPW are working on a smart metering data project. OCWD regularly shares data with agencies in the OC region which draw water from the aquifer it manages. This is an ongoing data sharing relationship where data about water quantity and quality is shared between OCWD and the agencies using its water. It should be noted that very little real-time data is shared. However, most agencies are collecting weather data and hydrological data, which can be shared in real time. The weather data are typically of high quality, but the hydrological data needs QC. Most data that is shared is reviewed and approved by agency employees before being shared. Additional data are often provided by agencies to requestors for such things as environmental site assessments. Most data are public.

A large amount of water quality data are collected at varying levels of sensitivity. There are many compliance obligations that necessitate collection of the data and all agencies have concerns that data may be misinterpreted if viewed by parties without expertise in the area. Participants indicated that it is very labor intensive/not streamlined to obtain and analyze data, which impedes decision making and that data generation in a timely manner is important. However, they are generally not reluctant to share data with other agencies. Another type of data that may be collected is AMI (advanced metering infrastructure / smart metering network), which is used to capture data at outflow sensors to come up with conclusions such as the correlation between irrigation at night and possibly reducing pollution.

Data are often shared via: ESRI's GIS-based platform, CEDEN (CA environmental data exchange network), dashboards, AWS CLI, uploads or emails, uploading to an FTP server. Most data is not real-time and is shared after QC. Agencies in this group did not have a problem sharing data with other agencies, but had concerns about sharing certain data with the public when that data may be misunderstood. They also indicated that they did not always have the technical staff to process and protect the data.

This group indicated that data sharing opportunities exist where jurisdictions overlap and when joint decisions must be made between agencies managing different components of a regional water system. A barrier to this is the technological limitation of smaller agencies with less resources that collect less data and have less ability to analyze and share that data. Agencies that could more easily use a platform such as SWADE, or have interest in this area, likely already have data tools they have developed.

Riverside County Agencies

Respondents indicated that different types of agencies have different needs/goals for data sharing. For example, water wholesalers who buy and resell data to other agencies might be

interested in more aggregated data versus agencies who deal with end consumers may be interested in meter-level data. Larger agencies that have member agencies share data with those agencies. That data are mostly shared with public agencies, including cities such as Chino Hills, Upland, and Ontario. Building relationships between agencies is key to increasing data sharing. Agencies with close relationships are more likely to share data.

Data are often shared via: NDA, CaDC platform, email, and excel. Data are often in different formats and participants indicated that consistent data representation would be useful. Data privacy was a concern for this group due to the fact that data may be misunderstood. Interpretation and context need to also be present when sharing data, especially with the public. Data format differences is also an obstacle to data sharing as data representation is not consistent and errors may propagate. Data sharing may also be reduced due to the time and labor required to format and share the data. The cost of managing a data platform was also a concern and agencies indicated that the costs may be better allocated elsewhere given the range of responsibilities of water agencies when there is no clear value or benefit to the agency in sharing the data. It must be made clear what purpose sharing the data serves. This group indicated that having access to other agencies' data would allow them to learn from those agencies experiences and tactics to better manage their own supply (especially in the context of outages and emergencies).

Key Findings

- Respondents indicated a need for an intuitive dashboard that is easy to upload and access data, with version control, and for which metadata and usage is well-documented.

- Respondents indicated a need for tools to help in decision making (e.g., water quality, management of pollution, assessing value of investments).

- The platform must be easy to operate for a non-data scientist.

- Agencies worry that data can be misused or misinterpreted by outsiders.

- QA/QC is a concern as data must be checked and approved by employees before being shared, impacting real-time platform functionality.

- Privacy and security issues are a concern for some agencies. Others indicate that it is less a concern where data is aggregated to protect individual's privacy.

- There is hesitation to share data when it may reflect poorly on an agency's performance, although regular operational data sharing among agencies during emergencies is useful.

- The water industry is historically risk averse. Using new technologies such as water data sharing platforms is resisted even though access to other agencies' data could allow learning from others' experiences and tactics.

- Understanding the challenges facing other agencies is useful to spur collaboration. Agencies share compliance data, water supply, production, quality data with research organizations, partner agencies or other agencies.

- Because there are already many data sharing efforts via ESRI's GIS-based platform, CEDEN (CA environmental data exchange network), dashboards, AWS CLI, uploads or emails, uploading to an FTP server, one drive, private cloud, thumb drives, NDAs or excel) respondents want to know how we can deepen our collaboration with these.

- Size of organization often dictates data usage and needs - this should be considered when designing a platform.

- Data sharing opportunities exist where jurisdictions overlap and when joint decisions must be made between agencies managing different components of a regional system.