ISSUE BRIEF

CIVIC ENGAGEMENT AND WATER DATA: HOW CAN CALIFORNIA MAKE DATA WORK FOR DECISION MAKERS?

Alida Cantor
Michael Kiparsky

September 2020

Berkeley Law | Center for Law, Energy, & the Environment

I. INTRODUCTION: CIVIC ENGAGEMENT IN CALIFORNIA WATER DATA

Much attention is focused on water data in California, and for good reason. California is struggling to make decisions that adapt its water systems to pressures such as climate change and population growth. Fully informed decisions are impossible without accurate and accessible data and information to support them. California already produces immense amounts of water-related data. However, agencies and stakeholders currently struggle to bring these data together to support routine decision making. Overall, we lack any systematic understanding of whether and how existing data can support the needs of decision makers.

California state agencies are working hard to plan and implement more effective water data systems in California, spurred in part by the Open and Transparent Water Data Act of 2016 (AB 1755).[1] The law, which charges state agencies with integrating water and environmental data, has presented a major opportunity for changing the way we think about, use, and improve our currently fragmented trove of water data. The nonprofit California Water Data Consortium recently launched, with the aim of fostering collaboration among experts and stakeholders during the implementation of AB 1755. [2] A primary focus of these initiatives is on increasing the usability of data for water-related decision making. However, whether and how California can improve data systems in a way that informs decision making may hinge upon civic engagement of water data users.

In this issue brief, we build upon the premise that civic engagement in water data—which we define here as participation of public, private, nonprofit, and community water data users and decision makers in issues around water data availability and data system design—is an important part of ensuring that data are useful and used to inform water decisions.[3] This issue brief is motivated by two key questions:

- What is the ideal role for civic engagement in the production of California's water data systems?
- How can we concretely enable civic engagement to inform investments in water data?

To address these questions, we interviewed water data experts from California and elsewhere.[4] Based on these interviews, we argue that civic engagement is crucial, but greater

clarity is needed about its most appropriate roles. In this paper, we compare and discuss several different models for fostering civic engagement in water data.

Who are water data users?

Water data users include a wide range of people and organizations such as:

- Governmental agencies and their staff (e.g., California Department of Water Resources and the State Water Resources Control Board)
- Public water agencies
- Community and environmental organizations
- Researchers and academics
- Members of the general public

Within and between each of these groups, there are many individuals with different levels of expertise, different needs for data, and different needs for analytical decision-support tools and visualizations.

II. EXPLORING DIFFERENT UNDERSTANDINGS OF CIVIC ENGAGEMENT IN THE WATER DATA SECTOR

All of the experts we interviewed articulated a deep commitment to the importance of civic engagement around water data. At the same time, common understanding of what that entails in practice is lacking. We describe two main schools of thought regarding civic engagement around water data, which we term an open data approach and a user centered approach. Each of these two approaches results in different strategies for civic engagement, including different strategies to engage decision makers and members of the public in data provision and data analytics (Figure 1). It is important to emphasize that our descriptions of these approaches are illustrative syntheses of a spectrum of stakeholder views, rather than formally articulated policy positions.

Open data approach

The primary focus of the open data approach is on making data available. The premise is that if data are accessible and transparent, developers, community members, and water data users will seize the opportunity to analyze it and develop the necessary tools to answer the questions they consider the most important. This follows from anecdotal accounts in other sectors where the availability of data spurs the emergence of novel information. In some cases, it does so through methods that were not foreseen before entrepreneurial innovation unlocked creative solutions and even surfaced problem statements that were not previously articulated.

Flexibility is a key advantage of the open data approach. Ideally, an open data approach facilitates developing a "data ecosystem that could do a thousand new things... rather than building one-off activities."[5] Open data can also support replicability as a basic scientific principle: "people need to have the data put out in its raw state in order to have confidence that they can repeat the analysis." In a recent example of successful implementation of the open data approach in the water arena, civic collaborators participated in a "datathon" and worked with the California State Water Resources Control Board to build a user-friendly interface for the public to easily check drinking water quality in their own hometown.[6]

The open data model also raises uncertainties. In concept, a creative and motivated individual or organization may volunteer to generate analytics and applications to support any given decision making need. However, this outcome is not guaranteed without a dedicated funding source, legal or regulatory requirement, or other incentive. As with any activity, success may hinge on whether funding, or the potential for future funding, is involved –in California relevant funding sources are ad hoc at best. Another problem is that an open data approach does not provide a means to prioritize either the provision or use of particular data among a wide range of potential sources. "The pitfall of this method is that it can be a black hole... there are all kinds of data that could be open, you can spend any amount of time and money just getting it out there." Moreover, data from numerous sources will have a highly variable range of standards and will not be uniform. And finally, it is not certain that such bottom-up efforts will comprehensively and equitably cover the wide range of needs that have, and have not yet, been identified.

Are water data different?

Making data free can spur innovation for social change. For example, in 2007 a computer programmer in Oakland took the initiative to generate an online map of crime data scraped from previously opaque city websites. The resulting transparency spurred real-world changes in police practices.[7]

This anecdote is one of many that illustrate the power of effective use of open data. But would this notion apply to water in the same way? To what extent does opening water data to the world result in systemic change in water decision making? There may be significant differences in water that should be considered.

Water is complex, fragmented, and niche. It encompasses a huge range of relevant entities and decisions. Useful metrics could be developed, particularly where there are policy-relevant decisions that are localized in nature or point to specific systemic failing. But given the importance of integration and complex synergies in next-generation water innovation, simple metrics will often be misleadingly simplistic.

To illustrate, crime rate and distribution can be clearly expressed and understood based on a simple graphic, and almost everyone in a city can understand its implications. Similarly, failures to provide sufficient water quality to disadvantaged communities have been mapped in simple and compelling ways[8] – such efforts would ideally motivate state-level policy change and spur more effective local management.

But the majority of water decisions are based on specialized and complex efforts. For example, water rights determinations are unlikely to be viewed as credible if based on an app generated solely by volunteers or even paid consultants. And efforts at the cutting edge of water innovation, such as those to develop managed aquifer recharge projects based on available floodwaters, are unlikely to be usefully represented without extensive effort to clarify specific contexts and incorporate detailed qualitative data sources. Further, there are thousands of decisions made on a daily basis to run the myriad water systems in the state – for how many of them will talented volunteer programmers step up to produce credible synthesis?

The message for decision makers is that data availability may well spur innovation, but it is far from certain that open data alone would fulfill the state's needs for data-driven management.

User centered approach

In contrast to the open data approach, a user centered approach focuses first on developing a more comprehensive understanding of data users' needs.[9] Then, data sets and decision support systems can be prioritized, developed according to user standards, and made available.

One strategy for implementing a user centered approach has been the development of 'use cases.' Use cases are short examinations of how water management decision processes employ data. The approach puts the data user front and center in the development process in order to assess data needs and communicate those needs to technical developers. A use case systematically addresses "who needs what data in what form to answer what question?" in order to guide effective data provision.[10]

The user centered approach counts as a strength its emphasis on making direct linkages between data and decision making processes, to enable "actual results for the world, ... which means informing a decision. Water decisions are hard questions, and if we don't answer those questions we are just having a good time with data." A user centered approach can also encourage efficiency by helping prioritize the most-used data sets for publication or inclusion in a system, and for the design of the system that federates and serves the data to users.

Limitations of the user centered approach include the need for upfront investment. First, public engagement can be slow and thus resource-intensive: "It takes a long time. If you're doing it right and truly listening to people and understanding what they need, you lose efficiency." Second, designing a data flow exactly as specified by data users may result in a product that is not only expensive, but inflexible, with limited ability to add new data or answer new questions. Certain data standards and harmonization strategies will also be prioritized to increase interoperability, which involves making analytical choices that may work for some users but not others. Third, prioritizing use cases can be seen as benefiting some interests over others, which may generate resistance: "You have to choose, and in that act of choosing, you prioritize particular users and data."

Figure 1: Civic engagement in water data takes place at multiple stages in the data life cycle (life cycle adapted from the NSF DataONE project). Civic engagement around data provision and data analytics takes on different forms in the open data and user centered approaches.

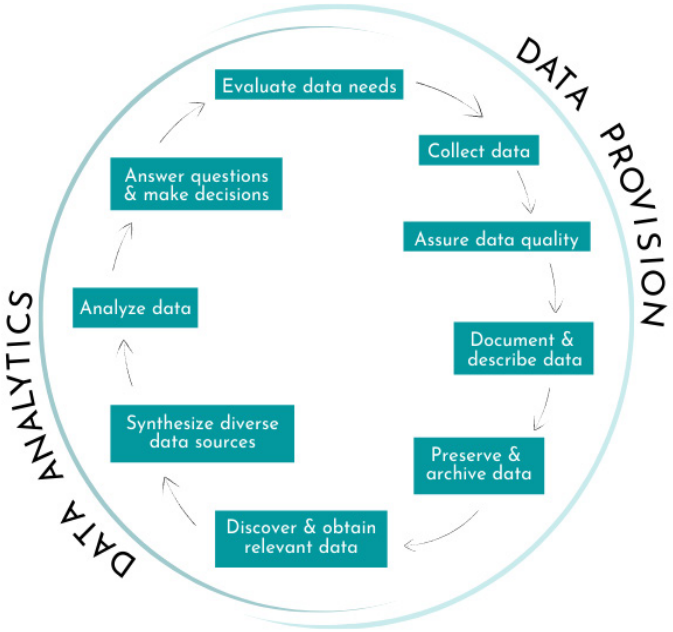

Table 1: Assumptions, advantages, and limitations of approaches to civic engagement in water data.

	OPEN DATA APPROACH	USER CENTERED APPROACH
	Shared assumptions: Civic engagement and usefulness of data are highly valued.	
Approaches to civic engagement	• Data provision: Focus on making the maximum amount of existing data publicly available. • Data analytics: Create and promote opportunities for public engagement in developing analytical tools, such as 'datathons.'	• Data provision: Focus on assessing data user needs and prioritize publication of most widely needed data. Use this understanding of data user needs to maximize the usability of shared data. • Data analytics: Through assessment of decision maker needs, focus on creating analytic tools that are determined to be most useful or needed.
Assumptions	• Teams of people will discover, develop, improve or refine open data resources to answer relevant questions. • Civic partners who understand community needs will bring their skills and resources to develop useful, accessible analytics and tools. • Efficiency comes from focusing on data provision and providing the flexibility for the public to work with data however they want.	• Starting with a user perspective helps data providers understand what data users need most, including parameters and interoperability concerns. • Case studies, or 'use cases,' can help develop a better understanding of data users' needs. Developing a subset of use cases will provide generalizable insight into which data sources and types are most needed. • Efficiency comes from ensuring that the most needed data are provided in forms that are most useful. Starting with data user needs helps prioritize so that the most widely-used data are provided first.
Advantages	• Flexibility and agility: a "data ecosystem" can be used for many different purposes. • Ability to respond to unanticipated challenges. • Focus on providing raw or 'low level' data ensures that analyses can be repeated. • Taps into "free" people resources and crowdsourcing to refine data and a range of potential final products.	• Builds in the assurance of relevance and usefulness: Data has a clear goal or outcome related directly to a problem. • Can help generate buy-in for water data systems because of focus on data users. • Focus on usefulness could help target resources for efficiency in a transparent way.
Limitations	• May not pan out in practice: developers may not step up to make the tools decision makers need. Assumes developers will be sufficiently incentivized to take action and develop tools and services. • Lacks a mechanism for prioritization of data or assurance of data quality. • Skills and resources to develop decision making tools are not always available to all relevant parties; analytics or tools may not be available to all.	• If use cases prioritize certain water users or issues over others, the resulting data systems may fail to address the needs of less powerful groups of data users. Broad participation is crucial. • Data resources could be inflexible if designed for a single purpose. Caution is required to avoid an overly narrow or inflexible design. • Given rapid changes (e.g., climate change) it can be difficult to predict which data will be significant in the future. • Stakeholder engagement can be resource intensive.

III. CHALLENGES, OPPORTUNITIES, AND DISAGREEMENTS AROUND INTEGRATING THE TWO APPROACHES TO INFORM OPEN AND TRANSPARENT WATER DATA

The open data and user centered approaches described above share an emphasis on civic engagement, but they each emphasize different aspects of data provision. Again, our descriptions of these approaches are syntheses of a spectrum of stakeholder views, rather than formally articulated policy positions. In addition, we find that not only do both of the approaches have merit, but neither is mutually exclusive. In the words of one expert: "So the answer in my book is really both, of course."

However, questions arise about how to operationalize these ideas, and how to take advantage of each. Each approach has limitations as well as advantages (see Table 1). Thinking carefully about how to blend the best elements of each approach to increase availability and usability of water data is a challenge for California's water data community. Here we explore some questions around civic engagement in water data, and discuss how each approach might inform the answers.

Public agencies, the main producers of water-sector data, inevitably work under resource constraints. Given these constraints, it is often necessary to make decisions around how to prioritize activities. For example, what data need to be collected, at what frequency and resolution? What metadata standards will be used? How will data be archived? What interfaces will best allow users to access data? What decision support systems or apps should be built to help data users answer questions?

The open data approach has several advantages when it comes to making the most of resources. Activities such as datathons can, in theory, cheaply harness public and volunteer efforts to create user interfaces and decision support tools and applications out of existing datasets. At the same time, even if adopting an open data approach, data sets and analytical tools must be prioritized in some way for quality assurance, quality control, publication, and ongoing updating and maintenance. A user centered approach can incorporate an understanding of data user needs to ensure that data with clear applications to real-world problems are prioritized, and that the process of prioritization is as transparent as possible.

In the sections that follow, we discuss use cases, a method for user centered civic engagement that has recently been applied in California. As the use case method is still in its nascent phases, definitive conclusions about its merits are not yet possible, but we examine some concerns and thoughts of water data experts.

How can use cases inform resource allocation decisions?

Use cases are one method of facilitating decisions about the prioritization of data and analytical tools. However, some respondents described a use case-based approach as 'political' to the extent that use cases themselves are subjective decisions that can implicitly guide the allocation of resources for data provision. Interviewees recognized that many important water-related issues are controversial and described tensions around wanting to stay out of 'hot topic' issues. When use cases focus on a controversial topic—for example, water transfers or curtailments based on limited water availability—"the concern is that use cases might muddy the waters between system capabilities and policy determinations." Relatedly, respondents were concerned that if a small set of specific use cases is developed, the selection of cases may have ramifications for system design, which may benefit particular constituencies: "you have to choose, and in that act of choosing, you prioritize particular users and data. And that's hard to do." That is, if certain use cases are used to inform data system development, other use cases are by definition *not* included,[11] which raises important questions around whose voices might not be heard during the process.[12]

Despite these concerns, interviewees also noted that an approach involving use cases can support a strategy for increasing equity and transparency: "There's a strong equity argument

for use cases, because they can make data available across the board more broadly." Without a structured process of evaluating data user needs, data prioritization and system design may be based on implicit assumptions. Use cases, on the other hand, can offer a mechanism for deliberation and prioritization of data needs, potentially increasing engagement for multiple constituencies and interests: "In a resource constrained environment where you have to prioritize, you need to have users guiding it to make sure it's actually valuable." If use case development includes a wide and diverse group of constituents, the process has the potential to be more open and transparent than approaches that rely on implicit prioritization. When conducted carefully and documented well, a process of developing use cases can lead to more open and transparent data systems.

How can use cases inform flexible, efficient, and useful data systems?

Respondents expressed concern that use cases take too long to develop: "there's thousands of use cases of water data. If you do the math about how long it takes to develop a use case, we'll all be dead if you do them all." Respondents also worried that a use case driven process could be inflexible: "When you take a use case and put it into the sausage making machine, the only thing the machine knows how to do is make that thing." In short, the approach is resource-intensive and may produce an end product that is not agile enough to meet changing user needs.

On the other hand, as discussed above, use cases can help prioritize open data efforts by identifying which of the many existing data sets are frequently used across different types of questions and topics. Then, the most frequently-used data sets can be prioritized in the process of making data open and transparent: "The data sets that come out of different use cases would pretty quickly be synergistic. You'd get more useful data faster than the kitchen sink approach." A data system that can be used to answer a very diverse set of questions should be able to also handle questions that have not yet been asked.[13] Ideally, a user centered approach should frequently re-evaluate user needs: a library of use cases should be a 'living document' that grows and changes over time, rather than a static set of information.

Water data involves equity considerations

Equity is an ongoing concern for California water generally, and is no less relevant for water data. Entities with sufficient resources (e.g., larger water utilities) can often independently create analytical tools and make use of available data. Meanwhile, many parties with pressing information needs (e.g., NGOs and disadvantaged communities) have fewer resources and may need external assistance to access and use data. If data are to reach all parties, then both open data and user centered efforts must include the needs of such under-resourced data users. Engaged and responsive government thus requires broader civic engagement around data: "For a government agency, any opportunity you have to sit down with users of data, there are other ancillary benefits of showing people you care. So, from a data perspective it's the right thing to do and from a good government perspective it's a good way to build capital in communities." Ideally, engagement around data can support a broader agenda of civic engagement, with a goal of ensuring that water data serves all Californians well.

IV. CONCLUSIONS: USING CIVIC ENGAGEMENT TO LINK DATA TO PROBLEM SOLVING

In conclusion, emerging concepts and developing paradigms around water data provision suggest some key takeaways for decision makers in state and local agencies, NGOs, and the private sector about civic engagement.

First, civic engagement is important for enabling efficiency, effectiveness, and equity in water data provision. Building avenues for diverse public input and involvement can beneficially

inform data provision. The open data and the user centered approaches distilled here each illustrate important strategies for civic engagement in water data.

Second, we recommend that those steering California's efforts on water data consider how to blend the positive elements of each approach toward civic engagement, and also that they remain wary of the limitations of a narrow focus on either in isolation. Data must be open, flexible, and transparent for unfettered public use, as the open data approach suggests. At the same time, the development of data and analytic tools must also be informed at multiple stages by the needs of data users. Evaluating user priorities in a structured way helps to develop a concrete understanding of decision makers' data needs. In many ways, the two approaches complement one another. For example, publishing less-processed data as emphasized by the open data approach is necessary for the success of a user centered approach; meanwhile, a user centered approach can inform which datasets and decision-support tools should be prioritized for development and open publication.

Third, we caution against the potential for short-termism in deciding how to weight investment in data provision. Specifically, a rush to publish data may be tempting for the sake of generating visible public milestones, but the state should analyze and articulate the extent to which doing so, without consideration of user needs, risks allocating resources by convenience rather than importance.

Fourth, data should ultimately aid problem solving. Achieving this outcome hinges upon both data availability as well as a consideration of user needs. "There has to be a user centered orientation. If it isn't user centered as a starting point, it will not generate buy in." The full potential of the current efforts around water data lies in increasing accessibility and usefulness of data, which necessitates engagement of data users throughout the data life cycle: "If AB 1755 just collects all this data and there's no use case for it or no way to access it easily then it's pointless. The value proposition is that different users can benefit by accessing the data, and to do that, it has to be structured and formatted and outputted in a way that it can be of service."

Ultimately, the entire enterprise of improving California's water data is an experiment in progress. We will not know until we try whether an open data approach will unlock innovation (whether widespread, commonly-used applications will be successfully developed through water datathons has yet to be seen). Likewise, we will not know until we try whether a user centered approach will ensure targeted and effective resource allocation (the idea that a sample of use cases can be extended to cover California's water needs, and that it will effectively drive improved water decision making practices, remains similarly untested). Until a clearer roadmap for the approach to data provision emerges, this uncertainty above all argues for an iterative portfolio of activities, including those based on both open data and user centered approaches.

ACKNOWLEDGEMENTS

We gratefully acknowledge the anonymous expert interviewees from the following organizations for sharing their perspectives and insights with us:

California Department of Water Resources
California State Water Resources Control Board
California Governor's Office of Planning and Research
Western States Water Council
Duke University
UC San Diego
Redstone Strategy Group
San Francisco Estuary Institute
Bechtel Foundation
FlowWest
Rosedale-Rio Bravo Water Storage District

We thank Mark Tompkins, Tony Hale, Joya Banerjee, Greg Gearhart, Eric Averett, Debbie Franco, Mike Myatt, and Niketa Kumar for reviewing the document and providing useful feedback that improved the issue brief. Rich Pauloo provided useful input. Cait Cady contributed research assistance. Funding was provided by the Water Foundation. Naturally, responsibility for all content, including any errors, rests with the authors alone.

ENDNOTES

1 AB 1755 directs state agencies to produce "a statewide integrated water data platform that, among other things, would integrate existing water and ecological data information from multiple databases."

2 California Water Data Consortium, https://cawaterdata.org/ (Accessed July 13, 2020). Mike Myatt, "Get to Know 6 Programs Driving Open Water Data," Water Foundation, July 9, 2019, https://waterfdn.org/what-we-are-saying/get-to-know-6-programs-driving-open-water-data/.

3 This issue brief builds on a previous report (Alida Cantor, et al., "Data for Water Decision Making: Informing the Implementation of California's Open and Transparent Water Data Act through Research and Engagement" (Berkeley, CA: Center for Law, Energy & the Environment, Berkeley Law, 2018), https://doi.org/10.15779/J28H01. This report developed the notion of user centered data provision, and a method for generating use cases to concretize data needs and inform the development of data systems. This issue brief reflects and extends these ideas, drawing on the 2019 report and subsequent interviews. See also Tara Moran, et al., "Evaluating the Use of Data Platforms for Water Management Decisions." Water in the West. Stanford Digital Repository. Available at: https://purl.stanford.edu/cb612zf3515.

4 We conducted interviews with 11 water data experts involved in AB 1755 implementation. Interviewees represented a range of relevant organizations from state agencies, NGOs, universities, and private sector organizations, as listed in the Acknowledgements. Interviewees remain anonymous per our research protocol.

5 All quotes in this document come from interviews conducted for this project. Based on our agreements with respondents, interviewees remain anonymous; we do not directly attribute any quotes used in this article by name or organization.

6 See the California Consumer Confidence Report (CCR) database, http://calwaterquality.com; for more examples of the open data projects, see the Datathon Resource Repository, https://cawaterboarddatacenter.github.io/Datathon-Resources/index.html.

7 Gavin Newsom and Lisa Dickey, Citizenville: How to Take the Town Square Digital and Reinvent Government (New York: Penguin Press, 2013).

8 See, for example, the California Consumer Confidence Report (CCR) database, http://calwaterquality.com.

9 See Cantor et al. (2018) for details on one possible process model for this approach.

10 See Cantor et al. (2018) for a more detailed template and guidance for use case design.

11 In our initial report (Cantor et al. 2018) we developed an initial set of 20 use cases chosen to represent a range of topics, with the explicit intent that these would seed the beginning of a larger repository rather than a complete or definitive set. This initial set is not meant to be definitive, but instead is meant to seed an open-source library to which additional contributions can be added over time.

12 For more on the topic of water data governance, see Nathan Huttner et al., "Governance and Funding for Open and Transparent Water Data" (Redstone Strategy Group, May 10, 2019), https://www.redstonestrategy.com/publications/ca-open-water-data/; as well as Cantor et al. (2018).

13 This articulation is aligned with a guiding principle of diminishing returns in use cases in computer science: if a system can handle a set of diverse questions or use cases, it can likely answer other questions that have not yet been asked.