Editorial

Where Are All the Data? The Case for a Comprehensive Water and Wastewater Utility Database

Christopher M. Chini, S.M.ASCE

Graduate Research Fellow, Dept. of Civil and Environmental Engineering, Univ. of Illinois at Urbana-Champaign, 205 N Mathews Ave., Urbana, IL 61801 (corresponding author). ORCID: http://orcid.org/0000-0002-1208-3646. E-mail: cchini2@illinois.edu

Ashlynn S. Stillwell, A.M.ASCE

Assistant Professor, Dept. of Civil and Environmental Engineering, Univ. of Illinois at Urbana-Champaign, 205 N Mathews Ave., Urbana, IL 61801.

DOI: 10.1061/(ASCE)WR.1943-5452.0000739

Motivation

A large majority of the U.S. population resides in urban environments, relying on water and wastewater utilities to deliver drinking water and remove wastewater from homes and businesses. These utilities record massive amounts of data, accounting for water quality, water volume, energy bills, and more. However, where are all these data? Currently the only means of obtaining these data in the United States are through open records requests to individual utilities. It is certainly beneficial to gain local insights into these data through open records, but this time-consuming process does not lend itself to national-level studies and is often lacking in response rates (Howard and McDermott 2016). The present survey recommends aggregate utility data were requested for water quantity and embedded energy requirements. In comparison with utilities in the energy sector, which report data to a centralized agency via the Energy Information Administration (EIA), no centralized database exists for national water and wastewater utilities. This significant data gap is highlighted and motivated through the challenges of contacting utilities across the country in an effort to create a representative national database, consisting of over 100 cities. The lack of this database creates research challenges and inhibits the benchmarking of utilities to further discussions of urban water consumption and the energy-water nexus. The implementation has significant benefits for university-utility research (Crow-Miller et al. 2016) and opportunities for advancing water resource sustainability and efficiency. This study calls for policy to create a national database for the purposes of water utility improvement and sustainability.

Where are all the data? This question is posed in an era of "big data," especially with respect to drinking water and wastewater utilities. Important research concepts, such as urban metabolism and the energy-water nexus, rely heavily on the availability and accessibility of aggregated empirical data, including water quantity and embedded energy. The EIA, which began in 1974 as a response to the energy crisis of the 1970s, provides a utility data model for the U.S. scale with respect to energy utilities. This example is not a perfect comparison to water utilities since energy utilities have a broader geographical reach, different operating costs, and staffing; however, it does show a willingness for policy to dictate data collection for the facilitation of comprehensive reports and studies. The deficit of water data presents a barrier to nationwide water research and limits studies to individual local utilities or regional studies.

Urban metabolism is the study of the flow and consumption of materials through a city or urban environment (Kennedy et al. 2011). With respect to water resources, urban metabolism studies estimate water flows comprise almost 90% of the total material flows through a city by mass (Wolman 1965; Decker et al. 2000; Kenway et al. 2011). However, the greatest challenge of assessing urban metabolism and, subsequently, urban water sustainability is the availability of data (Sahely et al. 2003). The energy-water nexus describes the interaction of water and energy resources, with water needed for thermoelectric power generation or fuel refinement and energy needed for water treatment and distribution. Water treatment for both potable water and wastewater requires a significant portion of U.S. electricity—approximately 4–16% of total electricity consumption, depending on the inclusion of water heating (Goldstein and Smith 2002; Twomey and Webber 2011; Sanders and Webber 2012). Therefore, no water utility data set would be complete without the cataloguing of energy consumption and, in the case of wastewater, energy recovery.

Presently, water resources have an extensive collection of publicly available data sets at the national level, just not on a utility scale. The USGS compiles two major data sets as part of the National Water Census: Water Data for the Nation and the National Water Use Information Program. The USGS Water Data for the Nation data set is an inventory of surface water and streamflow gauges across the country, with many properties having near real-time collection (USGS 2016). The National Water-Use Information Program is a county-level inventory of water resource withdrawals occurring every five years (USGS 2010). However, county-level inventories account only for water resources that are withdrawn from the given footprint; they do not account for intercounty water transfers such as those in the major metropolitan areas of New York or Los Angeles. The Clean Water Act and the Safe Drinking Water Act require annual water quality reports for drinking water utilities, which provide some data; however, these data are often limited to a snapshot of water quality and rarely include water volume or energy consumption. Some western states do collect flow data in a centralized database, such as the California State Water Resources Control Board (WRCB), but they do not provide energy characteristics of the system (WRCB 2016). Additionally, the American Water Works Association (AWWA) provides an annual benchmarking tool for water utilities, including aggregate energy consumption data at a utility level (AWWA 2015). However, these data are not freely available, nor does the tool provide data at the individual utility level.

Quest for Data

Open-records requests were sent to over 200 water and wastewater utilities in 112 cities representing all fifty states and the District of Columbia. Selected cities had a population greater than 100,000, except for states where no such city existed, in which case states' largest city(ies) were selected. In the current system, open-records requests provide the only means of utility-level data collection for

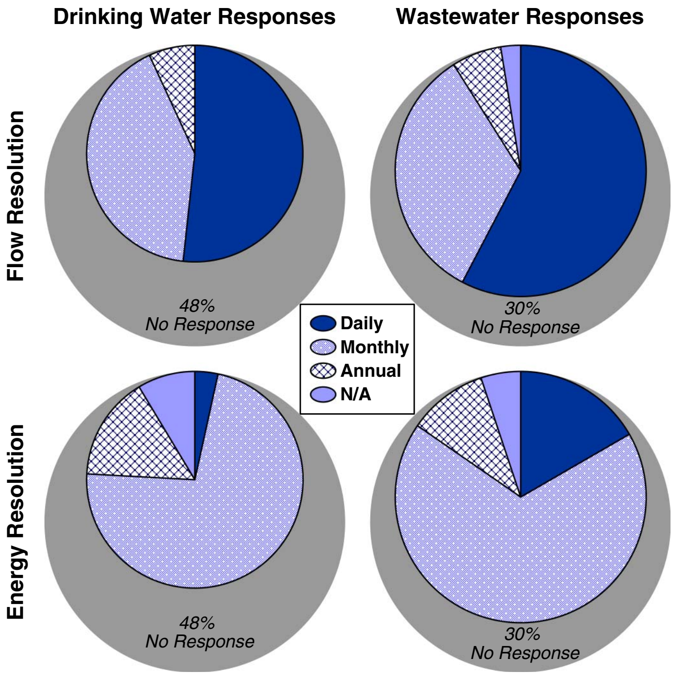

Fig. 1. Wastewater utility responses compared with drinking water utility responses; wastewater utilities provided more responses at a daily temporal resolution for both water and energy; water flow reporting is more likely than energy reporting to occur at the daily time scale; N/A = utility responses omitting either flow or energy data

researchers and the public. Data requests were for a single year (2012) and included daily operational data for treated flow, electricity consumption, natural gas consumption, and biogas generation. Although the current study did not request it, water quality data are critical component in future comprehensive databases. The reported statistics represent a 10-month data collection effort from October 2015 to July 2016. Utilities were contacted and requests made through various modes of communication, including standard mail, social media, phone calls, email, and online request forms. Of the contacted utilities, 97% were public and the remaining 3% were private (all drinking water). At the time of writing, 136 utilities responded to the data requests, for a response rate of 61%; no responses were received from private utilities. Wastewater utilities responded at a higher rate than drinking water utilities (Fig. 1).

Data Availability

Data availability references the issues that arise from a nonstandard data collection process. Personal communications with several facilities dealt with the ready availability of data from remote sites such as pump stations. Many utilities cited the large number of pump stations in remote locations, all operated on separate accounts and not within the direct purview of the contact, as a challenge in procuring the requested data, specifically energy data. Additionally, some utilities responded to data requests with nondigitized energy data documentation (e.g., copies of monthly energy bills), suggesting that energy bills are paid by the water utility's accounting department with no analysis by the water utility itself. Finally, the temporal resolution of energy and water data varies widely between utilities. Fig. 1 shows the varying temporal resolutions between water and energy data for both drinking water and wastewater services. The figure shows the reporting time step for both water and energy in a format promoting comparisons between the utility and data types. The discrepancy in time scales between water and energy data impedes decision-making opportunities in the energy-water nexus space.

Data Accessibility

In addition to data availability, the accessibility of data by researchers or other entities is a significant barrier to obtaining and creating a national database. Many utilities charged a fee for processing the data request. From the study, 6% of the utilties required a fee prior to data delivery, ranging from \$0.69 to more than \$500.00 for data that are relatively simple. Although this is a relatively low cost for individual utilities, the cost of assembling a comprehensive national database at these prices is significant for a single research group or project. Also, some utilities refused delivery of data without a nondisclosure agreement or simply declined the request because of sensitivity and potential security risks. A certain utility cited "the potential for revealing potential vulnerabilities to our water system" (personal communication). The number of utilities declining to respond is relatively small, 2%, but it poses a challenge to the development of a national database.

Path Forward

Based on the aforementioned challenges, we recommend that a strategy be developed for water resources data at the utility level that are similar to what is currently done for utilities in the energy sector (EIA 2016). Because of these challenges and the time-consuming nature of data collection, even at the scale requested, it is necessary for future research to create and maintain a comprehensive water resources database at the utility scale. A national database of utility-level water volumes, energy consumption, and water quality data is essential for the advancement of water utilities and their sustainable operations. While it is recognized that water and energy utilities are inherently different in geographical scope, operating costs, operating staff, and the like, comparing them represents an opportunity for understanding the differences.

Without this database, utilities lack a means of national comparison for promoting efficiency and driving toward sustainability. We call on policy to encourage, incentivize, and/or direct utilities to publish their data, as the EIA does with energy data. This policy could originate from nongovernmental organizations, such as the ASCE, national governmental agencies, such as the USGS or the Environmental Protection Agency (EPA), or research groups/universities. These entities have infrastructure in place that collect and maintain large databases on national levels, requiring minimal expansion of capacity to incorporate a new database. However, the database could also be implemented and maintained at the state level, expanding on existing water resources databases, such as those maintained by California's WRCB or the Texas Water Development Board and the Texas Commission on Environmental Quality. The creation of this database has significant potential to enable a wealth of future research and advancement toward sustainability, beneficial to water utilities across the country and around the world.

The details of reporting, requirements, and data quality control are reserved for subsequent steps in the policy-making process, specifically a policy implementation plan. As a starting point, it is proposed that this database include three categories of data for both drinking water and wastewater treatment facilities: quantity, quality, and process. Water quantity data would reflect volumes of treated water and billed water, facilitating discussions about water consumption and water loss. Water quality data would vary

J. Water Resour. Plann. Manage., 2017, 143(3): 01816005

between drinking water and wastewater facilities, but would include data on both inflow (raw) and outflow (treated) water resources. Process-level data would include type of treatment process, statistics on the utility and its network, energy required for treatment (including source, such as solar), and any generated energy from biogas. Additionally, the water quantity and quality components for wastewater would provide a convenient means of tracking combined or separate sewer overflows. Any policy should, at a minimum, require these statistics on a monthly time scale, but encourage the collection and reporting of all values on a daily basis.

The creation of policy for this database is justified through increased research and collaborative possibilities for water resources supply and treatment at the utility level. Benefits to utilities with this database include research and utility collaborations for studies on water loss, water conservation, energy for water, and learning about customer consumption habits. Additionally, these benefits equate to shared costs of research and the ability to complete larger analyses beyond the capacities of limited utility staff. From a research perspective, the database enables comparative sustainability analysis of water resources, such as a recent article comparing per capita drinking water consumption across the globe (Noiva et al. 2016). The implementation of the proposed water utility database has benefits for both water utilities and research institutions in enabling collaboration and comparability studies for the advancement of water resources sustainability.

References

AWWA (American Water Works Association). (2015). "Benchmarking performance indicators for water and wastewater utilities: 2013 survey data and analyses report." Denver.

Crow-Miller, B., Chang, H., Stoker, P., and Wentz, E. A. (2016). "Facilitating collaborative urban water management through university-utility cooperation." *Sustainable Cities Soc.*, 27, 475–483.

Decker, E. H., Elliott, S., Smith, F. A., Blake, D. R., and Rowland, F. S. (2000). "Energy and material flow through the urban ecosystem." *Annu. Rev. Energy Environ.*, 25(1), 685–740.

EIA (U.S. Energy Information Administration). (2016). ⟨http://www.eia.gov⟩ (Jul. 31, 2016).

Goldstein, R., and Smith, W. (2002). "Water and sustainability (volume 4): U.S. electricity consumption for water supply & treatment—The next half century." *Rep. No. 10067787*, Palo Alto, CA.

Howard, A. B., and McDermott, P. (2016). "Reforms to improve U.S. government accountability." *Science*, 353(6294), 35–36.

Kennedy, C., Pincetl, S., and Bunje, P. (2011). "The study of urban metabolism and its applications to urban planning and design." *Environ. Pollut.*, 159(8), 1965–1973.

Kenway, S., Lant, P., Priestley, A., and Daniels, P. (2011). "The connection between water and energy in cities: A review." *Water Sci. Technol.*, 63(9), 1983–1990.

Noiva, K., Fernández, J. E., and Wescoat, J. L. (2016). "Cluster analysis of urban water supply and demand: Toward large-scale comparative sustainability planning." *Sustainable Cities Soc*, 27, 484–496.

Sahely, H. R., Dudding, S., and Kennedy, C. A. (2003). "Estimating the urban metabolism of Canadian cities: Greater Toronto area case study." *Can. J. Civ. Eng.*, 30(2), 468–483.

Sanders, K. T., and Webber, M. E. (2012). "Evaluating the energy consumed for water use in the United States." *Environ. Res. Lett.*, 7(3), 034034.

Twomey, K. M., and Webber, M. E. (2011). "Evaluating the energy intensity of the U.S. public water system." *ASME 2011 5th Int. Conf. on Energy Sustainability*, ASME, New York, 1735–1748.

USGS. (2010). "Water use in the United States." ⟨http://water.usgs.gov/watuse/⟩ (Jul. 15, 2016).

USGS. (2016). "USGS current water data for the nation." ⟨http://waterdata.usgs.gov/nwis/rt⟩ (Jul. 15, 2016).

Wolman, A. (1965). "The metabolism of cities." *Sci. Am.*, 213(3), 178–190.

WRCB (Water Resources Control Board). (2016). ⟨http://www.waterboards.ca.gov/resources/data_databases/⟩ (Jul. 15, 2016).