

A New Database Offers Insights Into Municipal Drinking Water Systems

Sara Hughes and Christine J. Kirchhoff



Layout imagery by titima157/stock.adobe.com

Comprehensive understanding of US drinking water systems is challenged by the vast number and types of service providers and the lack of easily comparable data outside of drinking water system performance (e.g., population characteristics, environmental conditions). This is because data are often collected and housed by different groups at different scales and time intervals. For example, the US Environmental Protection Agency (EPA) collects information about water system performance from drinking water systems (e.g., source water, treatment violations) each month, while the US Census Bureau collects demographic information (e.g., income, race, education) useful for understanding the population served by drinking water systems at a different scale (e.g., block group or census tract) annually and every 10 years.

Environmental data, such as temperature and precipitation, that are useful for understanding environmental conditions that may affect water supplies are often collected at finer temporal scales but often over larger geographic scales. The lack of comprehensive drinking water data makes it difficult for researchers, decision makers, and practitioners to conduct meaningful analyses or identify important trends and patterns.

To help provide more integrated drinking water data, we compiled and made publicly available the Municipal Drinking Water Database (MDWD), which collates a range of data sources specific to US municipal drinking water systems. The database allows users to characterize the financial, institutional, demographic, and environmental conditions of municipal drinking water

systems over a period of 20 years. MDWD and all supporting information is hosted on the Harvard Dataverse and can be downloaded for free at <https://bit.ly/MunicipalDrinkingWaterDatabase>.

Creating MDWD

The key innovation in MDWD is linking drinking water utilities, or community water systems (CWSs), to the municipalities they serve. For this project, as a first step, we focused on CWSs that are owned and operated by a municipal government and serve at least 10,000 people. Making this link involved first downloading a full list of CWSs from EPA's Safe Drinking Water Information System and a full list of municipalities from the Census Bureau with 10,000 or more residents. We used a multi-step process of prefiltering, matching, and match-checking to generate a list of 2,219 CWSs and their corresponding municipalities. A more detailed description of the methodology behind this process can be found in our April 2023 *PLOS Water* article, "The Municipal Drinking Water Database."

Once these matches between CWSs and municipalities were made, we used the accompanying identification codes to integrate a range of additional data sets, including US Census and American Community Survey data, the Government Finance Database, Census of Governments, US Bureau of Labor Statistics, and the Voting and Election Science Team's political ideology data. Each data set provides additional insight into the status and operation of these systems and their communities. The final database has more than 30 variables available for each municipal drinking water system from 1997 to 2018, although some variables have coverage over longer time frames than others.

Using MDWD

MDWD can be used to ask questions, draw comparisons, and improve our understanding of CWSs in the United States like never before. For example, the database enables comparisons between drinking water systems, such as comparing the socioeconomic diversity of small and large systems. It's also possible to ask new questions, such as whether there are more groundwater-dependent systems in drier or wetter regions of the United States.

MDWD also provides opportunities to track trends over time, such as how demographics may be shifting for some systems or how Safe Drinking Water Act violations may correspond to subsequent changes in municipal government spending decisions. The demographic information contained in the database also allows users to more easily identify CWSs serving disadvantaged communities.

Because we developed a baseline match between key data sets, including geolocation information, MDWD users can incorporate additional data. For example, users could add variables from the US Census or American Community Survey; incorporate new environmental data sets; or bring in updated values from the US Bureau of Labor Statistics. Users may also find it useful to incorporate state- or location-specific data not available at a national scale but able to answer locally relevant questions. This might even involve citizen science projects with community groups in a given service area. The database can serve as a core resource on which to build a more tailored data set for a particular use or user.

The data and the insights they provide can also be useful for utilities and water industry leaders as they look to build relationships with other stakeholders and expand capabilities in their communities. MDWD allows utilities to easily identify other CWSs with similar demographic or financial profiles, providing a shortcut to fruitful collaboration or information-sharing. The database also allows individual utilities to compare or benchmark themselves relative to others across a range of performance metrics contained in the data set. As a publicly available and accessible database, MDWD could also be a useful tool for communicating and engaging with local schools and universities, helping students better understand their own drinking water utility as well as how it fits into a broader landscape of drinking water systems and environments.

Maximizing the Power of Data

Federal, state, and local decisions about investments and regulation should be informed by robust data and analyses. To date this has been challenging in the United States because of the fragmented nature of drinking water management and the siloed approach to data collection and storage. MDWD offers a giant leap forward, enabling better understanding of the conditions and trajectories of municipal drinking water systems and the communities they serve. The database provides the keystone in the collective effort to develop the data tools and resources to support efforts to ensure US drinking water is safe, affordable, and resilient to environmental change. 💧

Sara Hughes is an associate professor at the University of Michigan, Ann Arbor, Mich.; hughessm@umich.edu.

Christine J. Kirchhoff is an associate professor at Pennsylvania State University, University Park, Pa.; cxk475@psu.edu.

<https://doi.org/10.1002/awwa.2249>