(Mis)classification of water systems in the United States: Implications, analysis, and a proposed typology

Janice A. Beecher, Kyle Redican, Megan L. Kolioupoulos
June 1, 2020

Abstract
The national census and repository for water industry data in the United States is the Safe Drinking Water Information System (SDWIS) and widely used by researchers and regulators. Problems with the regulatory compliance data are well known, but the veracity of the data used to identify and differentiate water systems are equally important to the interpretation of research findings and the development of regulatory policy. Some reported findings may lead to simplistic and possibly misdirected recommendations. This research proposes a more exact classification scheme and typology, with a structural-function logic, and applies data mining methods and decision rules to code the 11,411 community water systems (CWSs) serving 73 million people in the Great Lakes region. We describe the challenge of establishing system identity and recoded a total of 536 misclassified systems. Our analysis highlights the striking asymmetry between the presence of governmental primary systems and nongovernmental ancillary systems, as well as the hybrid and interconnected nature of many water systems. We find that governmental systems are advantaged by system-level scale in terms of population served and water purchasing. For the long term, the ideal response is not to fix the data but to fix the system by which the data are generated for use. We provide policy recommendations for renovating SDWIS to improve system accountability, research credibility, and policy development.

Statement of Novelty
The Safe Drinking Water Information System (SDWIS) is a convenient source of information on all U.S. public water systems used by researchers and regulators. Problems with the regulatory compliance data are well known, but the veracity of the data used to identify and differentiate water systems are equally important to the interpretation of research findings as well as to regulatory policy development. This research proposes a more exact classification scheme and typology for water systems, accounting for both structure and function, and applies it to the Great Lakes region. We believe this study has potential to influence both research into and regulation of the water sector and urgency in the contemporary context.

Highlights
- Research on U.S. water systems rests on data lacking in validity and reliability.
- Federal data reflect the ambiguity of system types, misreporting, and miscoding.
- A comprehensive water system classification scheme and typology are proposed.

Keywords
Drinking water; regulation; Safe Drinking Water Information System (SDWIS); water system classification; water system ownership.

1.0 Introduction

A well-worn statistic in the U.S. water sector is the presence of an impressive number of nearly 50,000 community water systems (CWSs), employed rhetorically, liberally, and sometimes hyperbolically to rationalize regionalization, consolidation, and even privatization.[1] The number and the mythology around it belie the industry's nuanced structural and functional complexity. Too often, common generalizations and misunderstandings carry over to both research and regulation, with potential for misinformed decisions and unintended consequences. In this study, we take a closer look at the data and apply a comprehensive and refined structural-functional approach with an eye toward developing a more exact and useful water system classification scheme for census taking, needs assessment, oversight, and analytical purposes. Our "ground truth" rests on many years of applied experience as well as an iterative research process focusing on water systems in the Great Lakes region.

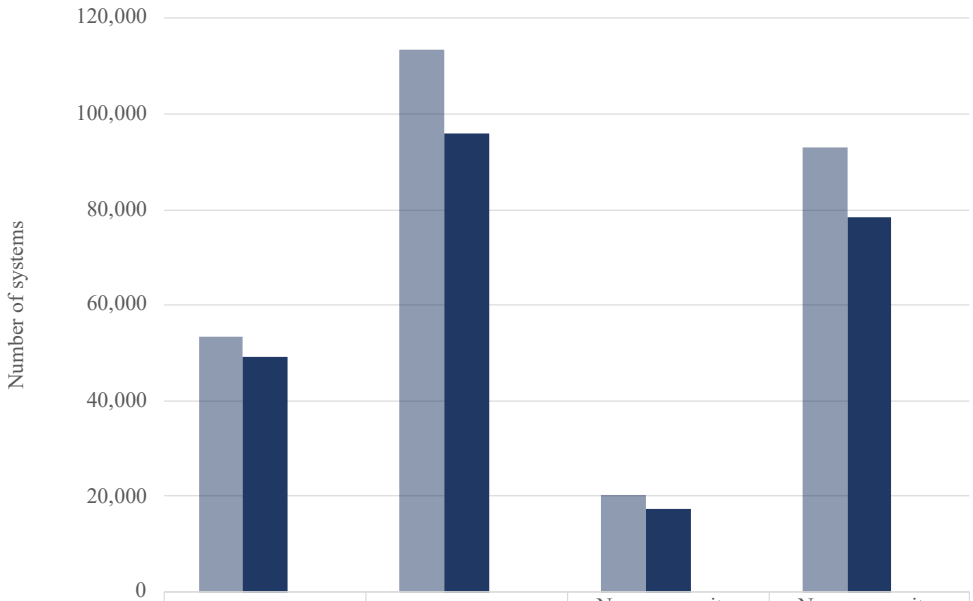
There is more to the U.S. water industry than meets the eye. The aggregate number of water systems can be misleading in several respects. The count can be accentuated by including the nation's nearly 96,000 non-transient and transient *noncommunity* water systems. We focus here only on community water systems, which are statutorily defined as public water systems that serve serves at least 15 service connections used by year-round residents of the area served by the system or regularly serves at least 25 year-round residents.

Dramatic graphics have been used to depict the industry and imply that all of the many CWSs are isolated, comparable, and static. In reality, the industry is dynamic, and water systems are shapeshifters: they expand, contract, and experience other forms of change relevant to regulatory policy (Rubin, 2013b). Trend data, summarized in Figure 1 indicate that the number of water systems has fallen steadily over recent decades due to net consolidation, likely influenced by both economic and policy forces. Over time, some systems also have morphed from one ownership form to another.[2]

Our map, accounting for system scale, is found in Figure 2 and shows that CWSs are unevenly constituted and distributed. The sector is fragmented but also bifurcated, with a relatively small proportion of larger systems serving the vast majority of the U.S. population. The numerous small systems are nonhomogeneous and pluralistic and may or may not be suitable candidates for structural or nonstructural alternatives based on locational and other variables.

[1] A prominent journalist recently opined on "two mostly unknown facts, both very important for public health. First, there are far too many US drinking water utilities. While there should be no more than several hundred for peak efficiency and getting the safest water, there are more than 50,000. (Not a typo!) Second, I reported that private drinking water utilities deliver significantly better health outcomes than do publicly owned municipal utilities," (Communication by Seth M. Siegel, January 6, 2019).

[2] One cooperative system in our study sample was recently authorized to become a local water district.



	Community water systems	Noncommunity systems	Noncommunity: nontransient populations	Noncommunity: transient populations
2000	53,410	113,375	20,334	93,041
2019	49,102	95,806	17,329	78,477
% change 2000-2019	-8.1%	-15.5%	-14.8%	-15.7%

Figure 1. Change in public water systems in the United States, 2000 to 2019.

Source: USEPA (CWSS, 2002) and (SDWIS (4Q2019). Includes tribal nation systems but excludes commonwealths and trust territories.

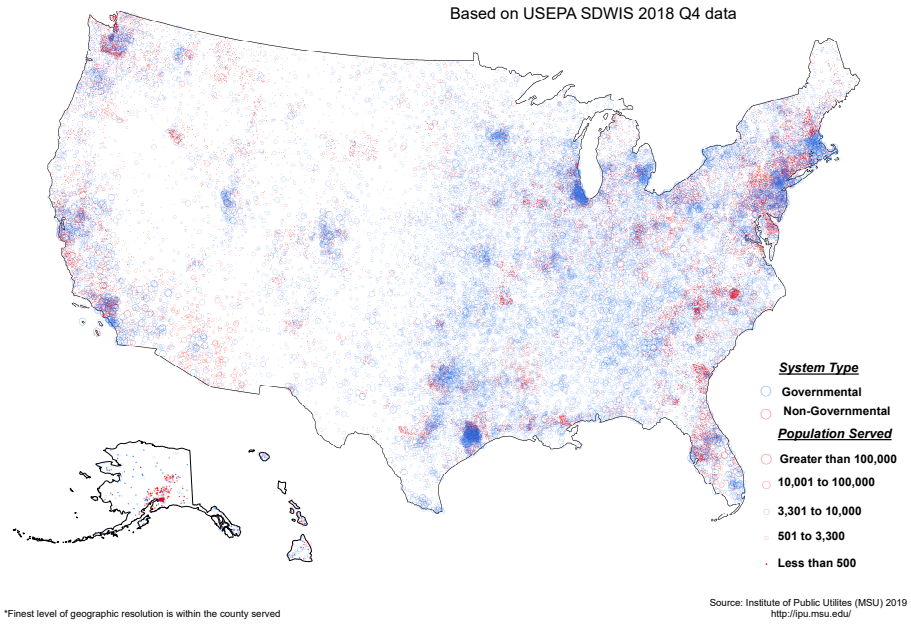


Figure 2. Map of U.S. water systems based on SDWIS accounting for system size.

The U.S. water sector also has a higher degree of connectivity than the total count implies. Many water systems have commonly shared owners, operators, and/or raw or finished water supplies. In other words, fewer systems are isolated than might appear. A substantial number of systems purchase treated water on a wholesale basis from another system, also referred to as "consecutive" systems (Grigg, 2018); about 19% of CWSs (serving about 23% of the U.S. population served by CWSs) is served by systems that purchase their water. These systems are not "ripe for consolidation" as they are already regionally structured to capture scale in production. Many systems share certified operators who manage multiple systems, particularly in rural areas. Some systems are operated within converged water and wastewater or water and energy entities.

Perhaps most importantly, drinking water regulators count water *systems,* not water *utilities.* Despite very different meanings, the terms are regularly conflated in the popular and even in the academic literature. A water system is a discrete *regulatory unit subject* to drinking water standards promulgated by the federal government and state primacy agencies.[3] A water system is a cohesive unit but can also consist of connected components, such as water wells. Adding to the confusion is that many CWSs do not actually serve communities in the conventional sense of the term, that is, local governmental units.

Utilities, not systems, are the regulatory unit from the perspective of the public service commissions and their numbers are not easily known. Not all water systems are operated as public utilities, but many utilities of different ownership structures operate multiple water systems. Many of larger publicly owned utilities provide water or operational services to other water systems. The larger private holding companies ("big water"), such as the American Water Works Company, own and operate hundreds of systems and may also operate other publicly or privately owned systems.

Smaller water systems are often challenged by disadvantages of scale and, in many cases, physical isolation and limited capacities. These systems are also highly pluralistic. Many smaller CWSs might have more in common with non-transient noncommunity water systems than CWSs serving large populations. A common conception is that structural consolidation of water systems will achieve economies of scale, but the most important of these are associated with production. Some economies in other areas can be achieved by structural but also nonstructural means (such as sharing and joint agreements).

Against this backdrop, the goal of this study is to delve into the SDWIS data to draw a more accurate picture of the U.S. drinking water sector. First, we briefly review some of the research that has relied on SDWIS data. Next, we detail our classification scheme for water systems and apply data mining methods and decision rules to code the water systems in the Great Lakes region. Finally, we provide policy recommendations for renovating SDWIS to improve system accountability, research credibility, and policy development.

[3] State standards must be at least as stringent as the federal standards. North Dakota is the only state that cedes primacy to the federal government. See https://www.epa.gov/dwcapacity/north-dakota.

2.0 Safe Drinking Water Information System (SDWIS)

The national census and repository for water industry data is the Safe Drinking Water Information System (SDWIS) of the U.S. Environmental Protection Agency (USEPA). Following the enactment of the Safe Drinking Water Act (SDWA) in 1974, the USEPA established the Federal Reporting Data System (FRDS) as part of its evolving program for Public Water Supply Supervision (PWSS). Hopes were high when it was transformed into the SDWIS in 1994, but its success was contingent on integration with state information systems (Tilton and Rosen, 1997).

The SDWIS database is updated quarterly based on submissions by the state drinking water primacy agencies, which they can choose to enter information through a Compliance Monitoring Data Portal.[4] The USEPA's periodic national surveys of CWSs and state financial needs follow the SDWIS rubric. SDWA compliance data are also incorporated in USEPA's Environmental and Compliance History Online (ECHO) and the Government Performance and Results Act (GPRA) tool.

The "SDWIS Fed Data Warehouse" includes (but is not limited to) the following searchable fields for each public water system, as reported by state primacy agencies:[5]

- Basic information, including: system name; identification number; city or county served; number of people served; type of system (residential, transient, non-transient); whether the system operates year-round or seasonally; and characteristics of the system's source(s) of water; number of facilities in the water system; county or city served; organization and administration names, with email, phone and fax numbers; the address of the company that owns the system.
- Violation information, including whether the system has: failed to follow established monitoring and reporting schedules; failed to comply with mandated treatment techniques; violated any maximum contaminant levels (MCLs); failed to communicate required information to their customers.
- Enforcement information, including actions states or EPA have taken to ensure that a public water system returns to compliance if it is in violation of a drinking water regulation.

Three key structural dimensions for water systems are size, ownership, and connectivity. SDWIS uses a consistent breakdown of size categories based on population served, a parameter relevant to operator certification. SDWIS also labels systems at one of six categories of ownership: local government, state government, federal government, private, public/private, and Native American.[6] SDWIS utilizes these categories to label systems as "public" or "private," when they are more accurately described as "governmental" or "nongovernmental," a material distinction. The "private" category

[4] See U.S. Environmental Protection Agency, epa.gov/enviro/sdwis-overview.
[5] See https://www.epa.gov/ground-water-and-drinking-water/safe-drinking-water-information-system-sdwis-federal-reporting
[6] According to USEPA SDWIS (4Q2019), 765 Native American systems serve more than 1 million people across the United States.

includes numerous not-for-profits systems and the public/private category appears to include systems for which ownership was not coded or unknown (not hybrid systems or public-private partnership, as the label might imply). Although the U.S. drinking water regulatory framework is agnostic as to ownership, utility ownership is potentially relevant to governance, incentives, and accountability, and thus performance.

The federal government evaluates compliance with the SDWA, as amended over the years, through programmatic reviews and SDWIS compliance data (USEPA-OIG, 2017). Timely, valid, and accurate reporting of compliance data are of obvious importance to meeting the statutory goal to ensure the safety of drinking water, yet problems with SDWIS and ECHO are well known and documented. USEPA notes "caveats" about reporting lag, underreporting, sample values, and missing or incorrect data, some of which are tagged with "alerts" to users that are both general and state-specific.[7]

Reports of the U.S. Government Accountability Office (GAO) in 2011 and USEPA's Office of the Inspector General (OIG) in 2017 found that some primacy agencies may have not been entering *monitoring and reporting* (M&R) violations as required (USEPA-OIG, 2017). M&R violations are troubling because they might mask exceedances of maximum-contaminant levels (MCLs) to the detriment of public health. The 2017 report concluded that the tools used by USEPA impeded consistent oversight, the agency was engaged in ongoing activities to improve data quality. Unfortunately, a decade-long effort to modernize compliance data collection and integration known as "SDWIS Prime" was put on hold in 2019.[8]

Researchers have noted inconsistencies between data reported in SDWIS and state sanitary surveys (Oxenford and Barrett, 2016). According to Josset et al. (2019, p. 433), despite "considerable progress has been made in data science and model development in recent years, data limitations continue to hamper analytics" and argue for the "development of a "national water portal." Based on our research, we also assert that the SDWIS data in their present form are not sufficiently valid and reliable for research purposes due to the ambiguity of system types as well as misreporting and miscoding. These issues are made worse by variability in state data management and oversight practices. Until the data are cleaned and verified (which requires considerable insight and effort), researchers should be more circumspect about drawing inferences about structure and performance and making related policy recommendations.

[7] See https://echo.epa.gov. "Primary Data Alert: Drinking water data shown in ECHO are based on violations reported by states to USEPA's Safe Drinking Water Information System. USEPA is aware of inaccuracies and underreporting of some data in this system. We are working with the states to improve the quality of the data. For more information, see the SDWIS home page" (https://echo.epa.gov/resources/echo-data/known-data-problems#sdwa).

[8] See https://e-enterprisefortheenvironment.net/our-projects/state-drinking-water-information-system-sdwis-prime/ and https://www.asdwa.org/2019/08/21/sdwis-prime-and-cmdp-update-8-21-2019. See also, "Some Inconvenient Truths About SDWIS by Greg Fabian at https://fabianpm.blog/.

3.0 Empirical Research Based on SDWIS

SDWIS is widely used to analyze the water sector, particularly with respect to compliance with federal drinking water regulations. A Google Scholar search for "Safe Drinking Water Information System" on January 1, 2020, yielded 715 findings in total, as well as 696 with "analysis," 680 with "population," 553 with "model," 531 with "compliance," 379 with "performance," and 370 with "ownership." SDWIS is enticing to researchers interested in the water sector precisely because it offers what appears to be convenient sets of both independent (explanatory) and dependent (explained) variables representing system characteristics. In some cases, data availability may drive the research more than the theoretical framing of the research question, at the risk of excluding factors that are likely pertinent but not readily available.

Quantitative researchers gravitate to SDWIS data to explore and model potential structural and other correlates of regulatory violations, parsed into health-related and management-related (such as monitoring and reporting and customer notification). Some of these studies are concerned more broadly with issues of fiscal health and environmental justice related to place.

The observed units are at different spatial scales. The system-level research (including Lee and Braden, 2008; Teodoro, 2014; Switzer and Teodoro, 2017; Allaire et al, 2018; and Teodoro et al, 2018) focuses on key water-system characteristics (such as location, population served, source water, and ownership). Some researchers have aggregated available data to the country scale (Wallsten and Kosec, 2008; McGavisk et al, 2013; Grooms, 2016; Greiner, 2016; Pennino et al, 2017; Allaire et al, 2018; McDonald and Jones, 2018; Montgomery et al, 2018; metropolitan statistical areas (Pape and Seo, 2015), census of governments units (Switzer and Teodoro, 2017), and zip codes (Marcillo and Krometis, 2019). SDWIS data have been combined with census data (Greiner, 2016; Switzer and Teodoro, 2017; McDonald and Jones, 2018) and data representing environmental conditions (Pennino et al, 2017), consumer spending patterns (Pape and Seo, 2015), and the urban/rurality of place (Marcillo and Krometis, 2019).

Although not meant to be exhaustive or parallel in statistical approach, some of the recent peer-reviewed research based on SDWIS data (including but not limited to multivariate modeling) finds the following, listed chronologically:

- "Policies that lower the political, regulatory, and physical cost of mergers may help make it a more effective tool to boost SDWA compliance" (Lee and Braden, 2007, p. 58).
- "The results suggest that absent competition, whether water systems are owned by private firms or governments may, on average, simply not matter much" (Wallsten and Kosec, 2008, p. 186).
- "Publicly owned PWS have slightly higher probability of violating MCL standards than privately owned systems" (Rahman et al., 2010, p. 264).
- "Smaller water systems are no more likely than larger systems, except very large systems, to violate health-related requirements… [and] appear more likely than

larger systems to violate monitoring, reporting, and notification requirements" (Rubin, 2013, p. E137).

- "The results indicate that income may play a factor in compliance, case studies also show that nonquantifiable factors have compliance implications." (McGavisk, Roberson, and Seidel, 2013, p. E115).
- "Utilities that are headed by professional engineers violate the SDWA significantly less frequently than do utilities led by nonengineers" (Teodoro, 2014, p. 983)
- "U.S. households are 25 percent more likely to purchase bottled water and increase their expenditures on bottled water 4–7 percent after news of a violation, a total change in expenditure of 28-32 percent." (Pape and Seo, 2015, p. 79).
- "Water utility privatization in the United States follows the logic of a variegated neoliberalism and constitutes a form of environmental injustice" (Greiner, 2016, p. 1).
- "There is basically the same probability of a violation in a system serving fewer than 500 as in a system serving up to 100,000" (Oxenford and Barrett, 2016, p. 35).
- "The relationship between compliance and availability of human capital is especially acute in smaller utilities, where limited organizational capacity means that organizations may struggle to attract and retain talented labor" (Switzer et al., 2016, p. E423).
- "Local engagement could affect compliance for both management and health regulations" (Teodoro and Switzer, 2018, p. 572).
- "[Ninety-five percent] of violations were from groundwater and the increase in proportion of PWSs in violation over time was from GW, not SW systems" (Pennino, Compton, and Leibowitz, 2017, p. 13453).
- "[Mobile home park] systems are more likely to incur more health-related violations than other systems" (Pierce and Gonzalez, 2017, p. 168).
- "In communities with higher populations of black and Hispanic individuals, SDWA health violations are more common" (Switzer and Teodoro, 2017, p. 45).
- "Private ownership and purchased water source are associated with compliance." (Allaire et al., 2017, p. 2078).
- "Our most consistent finding was the positive association of initial and repeat violations with the proportion of those who were uninsured, irrespective of stratification" (McDonald and Jones, 2018, p. 1401).
- "Using a data set of 168,823 municipal water systems in the United States from 2010 to 2014, we find evidence that stakeholder attention moderates the effect of ownership on compliance with drinking water quality standards" (Montgomery, Lyon, and Zhao, 2018, p. 207).
- "After committing a health violation, districts with low overall fiscal capacity are less likely to take out debt for water system investment than are nonviolating districts" (Scott et al., 2017, p. 12).
- "Private utilities were found to comply with both the health and management regulations of the SDWA at higher rates than public" (Teodoro et al., 2016, p. 52).

- "The incidence of MR violations is also significantly higher among privately owned than publicly owned systems" (Marcillo and Krometis, 2019, p. 6).
- "The associations we observed between nitrate concentrations and proportions of Hispanic residents support the need for improved efforts to assist vulnerable communities in addressing contamination and protecting source waters." (Schaider et al, 2019, p. 1).

Table 1 maps the variables extracted from SDWIS for inclusion in select statistical modeling in studies of various aspects of water utility performance. Different authors have different interpretations, but most include spatial (county) and scale (population) factors; some model at the enterprise (firm) level. Some include ownership and source water as explanatory variables. All of the models consider violation data, often as the dependent variable. Although in the aggregate, the rate of compliance with U.S. drinking water regulations is quite high (around 90-93%), many agree "that's not good enough" (Rose, 2019). Nonetheless, we note that the empirical research in this area is effectively modeling a highly skewed distribution of systems.

Table 1. Summary of SDWIS variables used in select multivariate statistical modeling.

	System type	Enterprise level	Aggregated level	Ownership type	Population served	Source water	County served	Zipcode	System age	Any violation	Health violations	M&R violations
Wallsten and Kosec (2008)	CWS			√	√	√	√			√	√	√
Lee and Braden (2008)	CWS	√		√	√	√	√			√	√	√
Mcgavisk et al. (2013)	CWS		√		√		√			√	√	√
Teodoro (2014)	CWS	√			√	√					√	√
Pape and Seo (2015)	CWS		√				√			√	√	√
Grooms (2016)	CWS		√				√			√	√	√
Pennino et al. (2017)	CWS		√	√	√	√	√			√	√	√
Switzer and Teodoro (2017)	CWS		√	√	√	√	√		√	√	√	√
Teodoro et al. (2018)	CWS		√	√	√	√	√			√	√	√
Montgomery et al. (2018)	PWS		√	√	√	√	√			√	√	√
McDonald and Jones (2018)	CWS		√		√		√			√	√	√
Allaire et al. (2018)	CWS	√	√	√	√	√	√			√	√	√
Marcillo and Krometis (2019)	CWS			√	√	√		√		√	√	√
Schaider et al, (2019)	CWS		√		√	√	√			√	√	√

Source: Authors' literature review. CWS = community water systems; PWS = all public water systems.

Not all of this research is sufficiently robust, theoretically or empirically, and it is unsurprising that the "evidence" overall is mixed and sometimes contradictory. The water industry is more complicated than even many experts realize, and several assertions in the literature raise red flags. Some studies (and the popular press) confuse CWSs (in one case, noncommunity systems) with water utilities. Except for controlling for primary water source, the research largely ignores how water systems are connected even though

many small systems are owned or operated by larger entities, including very large utilities.[9] Excluding small systems from a sample is a potentially significant source of bias regarding scale effects given these forms of consolidation. Other studies applied measures of firm concentration, despite the monopolistic character of the industry. Some studies seem to assume that systems labeled as "private" are corporately owned when this often is not the case. Privately owned water systems are especially heterogeneous and misunderstood; however, all of these are profit-oriented firms owned by one or more investors of their own capital who are entitled the opportunity to returns on that investment.

Modeling regulatory compliance or other aspects of performance is especially challenging in terms of the spatial and scale dimension and unit matching. Water systems do not have consistent local signatures. Service areas are not well specified in the available data and doing so would be very challenging. One provider may serve multiple geopolitical areas or multiple providers may serve one geopolitical area. It is common for larger municipal systems to serve outside of municipal corporate boundaries. It is also common for larger privately owned utilities to own and/or operate systems across communities and, in the case of holding companies, across states.

This research is only as good as the data and models on which it relies. The SDWIS data are limiting in terms of available parameters as well as data quality. Given the heterogeneity of the water sector, models must be specified to control for potentially relevant social, economic, fiscal, and political factors. Given weaknesses in the raw data, as reported in SDWIS, some reported research findings may be spurious, biased, and misleading, leading to simplistic and possibly misdirected recommendations.

Moreover, incidental findings may lack firm theoretical support or well-formed hypotheses. Yet causalities inferred from this research can be (and have been) appropriated and repeated in political contexts, particularly when a finding aligns with special interest.[10] Of course, correlation does not prove causality, and even correlation cannot dictate public policy. The need for better data and modeling has urgency in the context of contemporary policy and politics related to structural reform (including consolidation and privatization).

---

[9] The larger holding companies, such as American Water Works and Essential Utilities (formerly Aqua America), serve several hundreds of systems across states, some of which are actually very small. American owns five very small proximate but physically separate systems in Michigan's Upper Peninsula.

[10] Private interests are quick to offer purported "evidence" of superiority over the public sector. The head of a national trade group reportedly quoted "from an American Journal of Political Science article asserting the theory that 'public agencies tend to underprice their services and produce inferior quality relative to private firms'" (Harber, 2019). See also, National Association of Water Companies, "Studies Confirm Water Companies' Exceptional Record of Delivering Highest-Quality Water," available at truthfromthetap.com, and Seth M. Siegel "Warren Has a Plan for Your Drinking Water—but It's Risky," (Wall Street Journal (January 2, 2019).

4.0 Structural-functional Classification of Water Systems

Although less discussed than the compliance data, others have also recognized the need to improve understanding of industry structure. Rubin (2013b) tapped SDWIS to assess net changes in water industry structure, concluding that the data as reported are insufficient for drawing conclusions about the nature of consolidation We agree, although would also assert that these data still offers several important clues about industry structure, including various means of system connectivity. The issue is whether they are sufficiently valid, reliable, and adequate or could become so with better means of reporting.

In the context of a cross-national comparison, Grigg (2018) recently called for a new classification scheme focusing on access to and ownership of systems, and the possible relevance of water provider and water user perspectives. While we do not adopt Grigg's (2018) approach, we share his concerns and observation that "there is much to learn about ownership, governance, and organization and how they affect the quality of piped water services for billions of people."

In sum, we see considerable room for improvement in the aggregation of systems for analytical and policymaking purposes. Whether our proposed system would yield statistically significant differences in findings cannot be known without further research. Until some effort is devoted to SDWIS, caution is urged against overly broad and speculative conclusions, particularly with respect to guiding public policy.

5.0 Proposed Classification Scheme

To be useful, water system classification should be logical, conceptually clear, and comprehensive of all water systems. Identified categories should be intuitive, unambiguous, and mutually exclusive. In the ideal, each system should have an obvious and exclusive classification, and not depend on coder judgment. In practice, given the nature of many smaller water systems, this can be easier said than done.

Our scheme emphasizes conceptualizing the structural-functional nature of water system enterprises. It is informed by research experience in the sector, the applied literature, and insight that can only be gained by a granular examination of the data. We approached the challenge inductively using the identifiers available in SDWIS and available supplemental sources, which vary by jurisdiction. To classify systems, we used an iterative process to develop a set of decision rules for coding and revisited the classification scheme numerous times. Researching individual systems as necessary, we were able to classify systems using processes of judgment and elimination.

Key to the classification of water systems is establishing the accountable ownership of system assets and ultimate responsibility for system operations, performance, and regulatory compliance (that is, legally recognized owners of record). We utilize a system organizational chart (Figure 3 with details and definitions provided in Appendices A and

B) to present an intuitive four-level hierarchy that stratifies system categories and identifies specific types based on the underlying structural-functional logic of our classification scheme. We start with two foundational dimensions by subdividing CWSs into governmental and nongovernmental and then into primary and ancillary systems. This follows SDWIS generally, and the emphasis on ownership in the literature, but elucidates the asymmetry of structural roles in the sector, despite parallels across certain functions.

We stress that this ordering can readily be flipped (this is, the primary vs. ancillary dimension could be followed by the governmental vs. nongovernmental dimension). Our classification is deliberately configured to allow for different permutations of the types to support alternative research theories and hypotheses. In other words, because the types are meant to be mutually exclusive, the data can be "sliced and diced" to expand or contract categories and include, exclude, or recombine system types according to research interests and criteria.

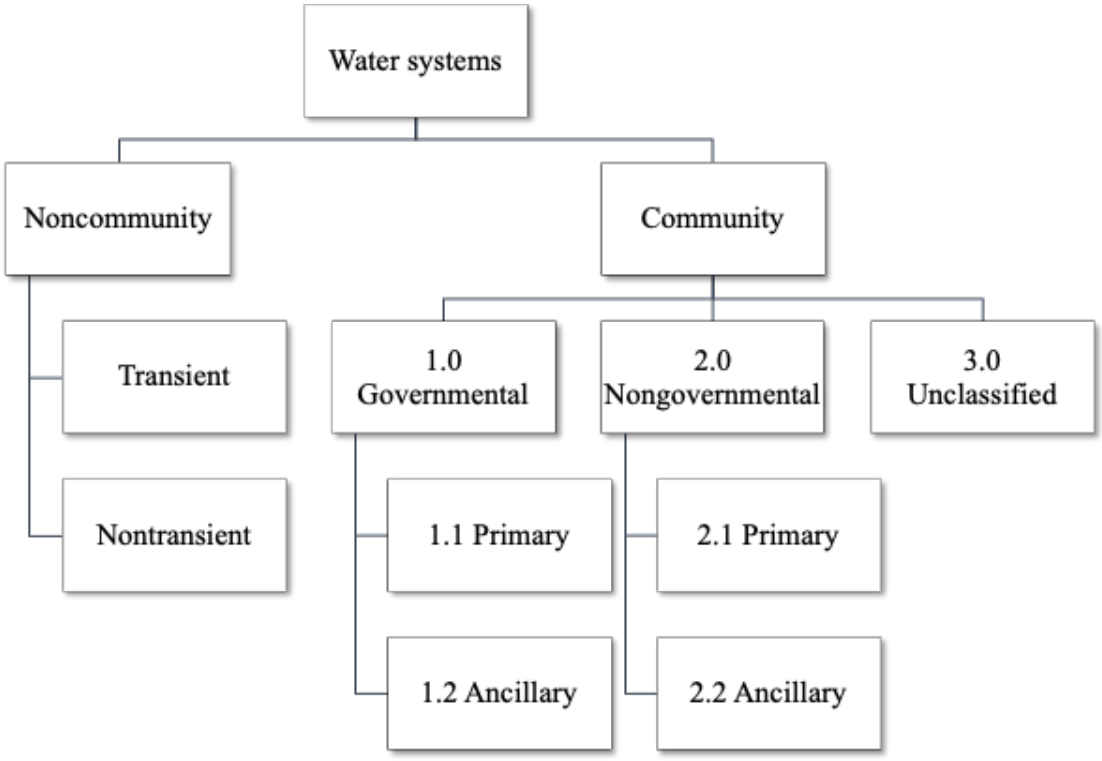


Figure 3. Water system organizational chart (levels 1 and 2).

6.0 Water System Typology

The goal of this research was to establish a set of mutually exclusive water system types and we arrived at a total of 40 (not including unclassified), detailed in Appendices A and B. With further research across regions not included in the scope of this analysis, some types might be redefined and others might emerge.

6.1 Community vs. Noncommunity Systems

CWSs are appropriately separated from the thousands of non-community systems that provide water to transient and non-transient populations. Separately, private water wells and septic systems are also prevalent throughout the United States. These distinctions matter to drinking water regulation and other policies. Our classification analysis focuses only on statutory CWSs.

6.2 Governmental vs. Nongovernmental

CWSs can be subdivided into governmental and nongovernmental systems. This relabeling corrects for the mischaracterization of systems by SDWIS as "public" and "private," respectively. The available information from SDWIS and elsewhere, while not infallible, are generally sufficient to make this initial determination.

The governmental category consists only of authorized and mostly autonomous units or subunits of the federal, state, local, and tribal governments.[11] Only the Navajo nation exerts primacy, while primacy for other Native American systems falls to the federal or state governments. Many if not most governmental systems also have taxing authority and other sources of revenue that could be applied to water infrastructure. Local governments (that is, municipalities) dominate the water sector in terms of population served, both nationally and in our sample. However, the data are insufficient to establish the details of a governmental system's enterprise status, including governance and fiscal autonomy. A municipal water system may be housed as a department of the local government or it may have considerable enterprise autonomy; some are governed by municipal boards or commissions.

The nongovernmental category is highly heterogenous, as elaborated in the further subcategories Although we identified several systems that are operated contractually, we did not attempt to code for this variable for all systems, even though this would certainly be of interest to researchers and policymakers and could easily be self-reported by systems going forward.

[11] The U.S. is home to more than 800 tribal nation systems serving more than 1 million people.

6.3 Primary vs. Ancillary Systems

Once governmental or nongovernmental ownership status is established, the system's structural and functional character defines whether it is a primary or ancillary system. The distinction between primary and ancillary water systems is acknowledged (Grigg, 2018) but not necessarily well understood. A novelty of our approach is that we draw this distinction within both the governmental and nongovernmental classifications. Functionally, primary systems distribute water as "utilities," whereas ancillary systems provide water to their "facilities."

6.4 Types of Primary Systems

A primary system is an entity whose primary function is providing water as a public utility service, a term that defines function but not ownership (that is, a public utility can be owned by a public, private, or not-for-profit entity). Utilities are understood "common carriers" that manage infrastructure and provide service on a compensatory basis to members of the public at large. Compared to a system, a utility is a "business unit" and its financial resources (including profits, if applicable) are derived from the provision of water service. Water used is typically metered and billed to residential, commercial, and industrial customers.

Primary governmental systems operate as units or divisions of government or public corporations; they are understood as publicly owned utilities (regardless of the number of systems operated or whether service is confined to municipal corporate boundaries). Included are general-purpose and special-purpose governments, the latter of which includes districts and authorities providing treated water on a wholesale or retail basis. We hold a place for public shareholder systems that are formally structured as wholly owned governmental corporations.[12] We also separately classify tribal nation systems, although without further refinement by function in this iteration. Tribal systems are not unlike other community water systems serving residential and nonresidential customers, but they can also provide service to health, educational, and recreational facilities. Tribal systems may fall under federal, state, or tribal primacy for regulatory purposes.[13]

Primary nongovernmental systems operate as either private for-profit or not-for-profit entities; they too function as public utilities. For-profit entities include publicly traded and private equity companies (domestic and foreign) that hold systems across states as well as independent companies operating intrastate to provide retail and sometimes wholesale services as well.[14] Due to their monopoly status and profit orientation, privately owned systems (or their parent utilities) are generally subject to state economic regulation of pricing; they are also subject to federal financial regulation if ownership

[12] Exemplified by Louisville Water Company, Kentucky, and the Pennichuck Water Corporation, New Hampshire. A deliberate method for differentiating these systems from other municipal systems is needed.
[13] At this time, only the Navajo Nation has primacy. The federal government, through its regional USEPA offices, has primacy for all tribal nation systems in the Great Lakes region, except one for which New York has primacy.
[14] Private equity is a distinctive for-profit ownership form that can be difficult to recognize.

shares are publicly traded. Not-for-profit systems include owner cooperatives (or mutual companies) and associations that have comparable structural and operating characteristics suggestive of utilities; we recognize, however, that associations can also resemble nongovernmental ancillary systems.

The particular water system types are based on the details of both structure and function. In each case, we center on the organizational character of the CWS enterprise and the residential population it *primarily* serves, allowing for diversity. This is particularly important for parsing systems in the smaller categories, which in some cases might be included in more than type.

6.5 Types of Ancillary Systems

For ancillary systems, water service is a secondary to a primary activity that is not understood as a public utility service, that is, the enterprise's financial resources are derived from something other than the provision of water service. USEPA has defined an ancillary system as one for which "providing potable water is not the entity's primary business… Instead, these systems provide water as an ancillary function of their principal business or enterprise. Examples include mobile home parks, schools, hospitals, prisons, and commercial businesses" (USEPA, 1997 CWSS, p. 6).

Ancillary systems have varied structural features, but after accounting for their governmental or nongovernmental character, we focused on their function over their form.[15] These systems tend to serve relatively smaller populations in more physically confined areas. Mobile-home developments may be the exception, but many ancillary systems might be structurally and functionally closer to non-transient noncommunity systems for regulatory policy and research purposes. Many of these systems may have little in the way of a water distribution network, which tends to define utility enterprises. Very small systems may have little if any infrastructure other than a water source (a water well or connection to another system) and the premise plumbing needed to serve a building or residential complex. In many cases, water resales (retail sales of water purchased wholesale) or even submetering may have created a regulated system. These realities also suggest that solutions to noncompliance must be tailored.

Governmental ancillary systems operate as units or divisions of government or public corporations whose primary function is not water service but that provide water to a residential population. They include federal and state facilities functioning in the military, correctional, healthcare, education and research, and parks and forests. Local facilities include public housing authorities (but not military).

Nongovernmental ancillary systems are owned and operated by entities whose primary function is not water service but that provide water to a residential population (including residential housing and property developments) or residential facilities that may be

[15] For example, a camp facility owned by a religious group would be classified as recreational; a residential facility providing supportive medical services would be classified as healthcare; a monastery would be classified as religious.

independent or affiliated with other institutions. Developer systems resemble nongovernmental primary systems, but owners are primarily in the real estate business (comparable mobile home parks), rather than the utility business; some are eventually turned over to a primary system of some kind (governmental or nongovernmental). Nongovernmental residential facilities have parallels to governmental facilities.

7.0 Coding Methodology

We tested our classification by applying it to eight states bordering the Great Lakes: Illinois, Indiana, Michigan, Minnesota, New York, Ohio, Pennsylvania, Wisconsin. Although our sample is universal for this region, it is not meant to be representative of the country. To an important extent, state political cultures, legal frameworks, and histories shape the prevailing structures of CWSs and thus comparability across states.

The sample provides some interesting stratification across system types and regulatory frameworks. In three states (MI, MN, and WI) the private sector has a very limited presence; in two of these (MI and MN), no systems are subject to economic regulation by the state public service commission (PSC). In Wisconsin, all municipalities are subject to PSC regulation; commission regulation applies to many of Indiana's municipalities, districts, and not-for-profit systems as well.

Our data mining and coding methodology involved a detailed and labor-intensive process by which systems were identified, coded, and sometimes recoded in a process that coincided with the iterative development of the typology. SDWIS reform is essential so that individual researchers do not need to expend this level of effort, which would be impractical and consume resources better devoted to analysis.

The steps in our coding and classification process are summarized in Figure 4 and detailed in Appendix C; each step in the process was revisited as necessary. The coding experience itself allows for sharing several impressions and lessons learned from recurring issues and frustrations. In particular, system names can be informative, especially for some system types (such as mobile home park systems), but not reliable. Vagueness and ambiguity are common. For example, one system whose name included "Subdivision Water District, Inc." was verified as a cooperative system.

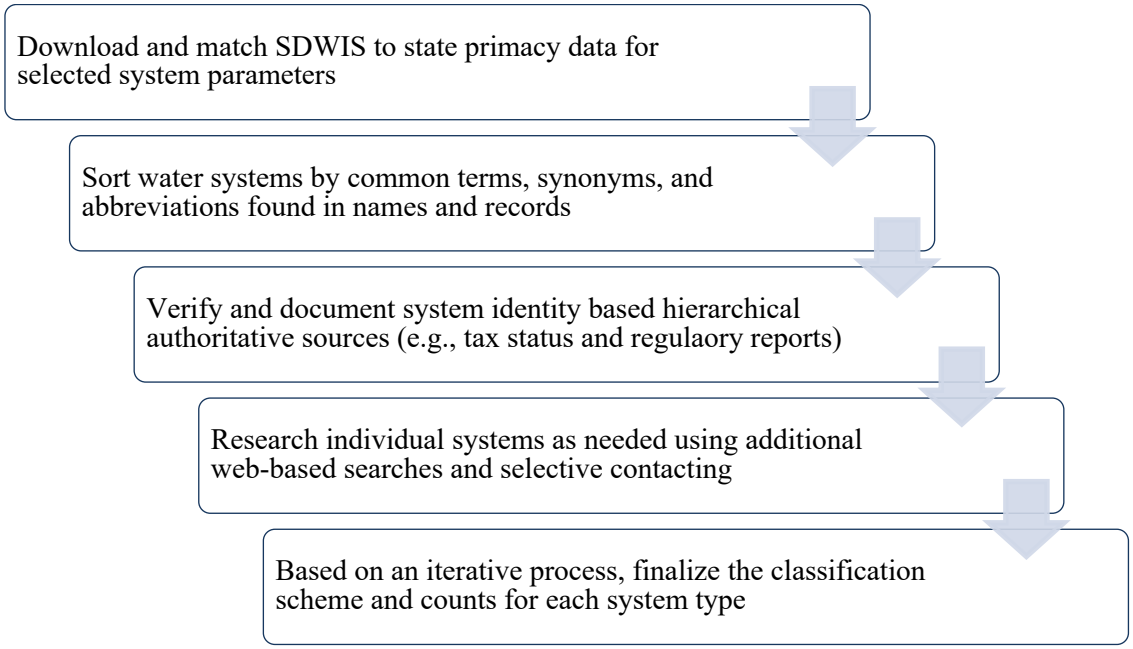


Figure 4. Summary of the water system coding process.

Other identity challenges associated with naming include the following:

- System names are not consistently used, with variations in abbreviation and hyphenation that made the validation process challenging.
- "Waterworks," "water department," and "water company" are names that may apply systems that are publicly, privately, or member-owned.
- Governmental entities may describe themselves as not-for-profit entities (e.g., "companies").
- Corporate identities and parents or affiliates are unclear and inconsistent, including across holding companies.
- Legacy names sometimes remain on systems following their acquisition by another entity.
- A "farm" may be a working farm with residential facilities or (more likely) a real estate development or subdivision that was previously a farm.
- A "golf club" is likely a real estate development associated with a golf club.
- A "park" may be a mobile home park, a real estate development, or an actual park.
- An association could be a homeowners' association or a member-owned cooperative.
- Universities and other educational facilities may be served by primary or ancillary systems of different ownership forms.
- Real estate developments and subdivisions are especially ambiguous and go by several names (including estates, commons, court, community).

- Residential developments may be single-family, multi-family, condominium, mobile or manufactured home communities and served by owner associations, cooperatives, or private entities.
- Systems managed under contractual arrangements or partnerships cannot be determined with any degree of accuracy.
- Names can be duplicated, as in the two systems named “Twin Lakes” in our sample.

Based on the available data and our coding methodology, we are reasonably confident about classifying the systems in the Great Lakes states according to our typology. Classifying systems according to structural and functional specifics at the third and fourth levels was more challenging, particularly for some types (including those involving various types of residential developments with ambiguous identities and lack of verification). Our typology also preserves some types for which few systems qualify based on statutory criteria. We use a placeholder for the public shareholder type, even though none are found in our Great Lakes sample, as well as for domestic private equity.

8.0 Applying the Classification Scheme

Great Lakes sample of 11,411 CWSs is broadly comparable to the national SDWIS sample of 49,648 systems, as shown in Appendix D. In order of the number of systems by state, the total systems and populations served are New York (2,292 and 18.2 million), Pennsylvania (1,940 and 11.4 million), Illinois (1,757 and 12.0 million), Michigan (1,380 and 7.3 million), Ohio (1,173 and 10.3 million), Wisconsin (1,044 and 4.1 million), Minnesota (965 and 4.5 million), Indiana (779 and 4.9 million), and tribal nation systems (82 and 123,323).

We note slightly fewer systems in the smallest population category (<500) in the region, although both New York and Pennsylvania have large numbers of systems serving populations less than 500 (1,427 and 1,124, respectively). Very large systems (>100,000) have more presence in New York (17), Pennsylvania (16), Ohio (12), and Illinois (10).

Recoding to correct for classification varied across states and system types. The largest number of changes were for systems in Indiana and Ohio, affecting more than 10% of their systems and representing more 20% of the population served. For Illinois, Michigan, Minnesota, and Wisconsin, the recoded systems represented less than 1% of their total population served. Pennsylvania was at about 6% of systems moved while New York was close to 4% of systems moved.

In Table 2, following the general SDIWS categories, we compare the data as reported by SDWIS and as recoded by our team. The net effect was to reduce the number of “private” systems by 42 and to eliminate all 117 systems tagged in SDWIS as “public/private”; these systems account for populations served of 935,072 and 1,905,567, respectively. Net gains are seen in the governmental systems, particularly at the local level, but 14 systems

could not be classified. SDWIS thus may tend to overestimate the number of private systems, while underestimating the number of local government systems

In Table 3, following the same general categories, we traced systems as reported (rows) to where they were effectively recoded (columns) as informed by our typology. We recoded a total of 536 misclassified systems, accounting for 4.7% of the systems in our Great Lakes sample and 7.5% of the population served (about 5.4 million people). Notably, a total of 301 systems were moved mainly from the private and public/private categories to local government, whereas 214 systems were moved mainly from local government and public/private to private. We consider these findings to be potentially significant in terms of performance modeling for the sector.

In terms of some specific types, a total of 140 local governmental retail-service districts and authorities were moved from various SDWIS designations (federal, state, private, and public/private). We moved 70 systems into the municipal (governmental) type and 72 systems into the cooperative (nongovernmental) type from other SDWIS classifications. We moved 11 private and two local government systems into a category for nongovernmental not-for-profit trusts. We found that all of the systems identified with tribal nations were correctly classified generally in SDWIS.

A few anecdotal misclassifications are material and noteworthy:

- The Cincinnati and Dayton public water systems (serving populations of 750,200 and 140,000, respectively), among others in the state, are misclassified as private systems.
- Three systems owned and operated by the Citizens Utilities not-for-profit public trust in Indiana, serving a total population greater than 870,000, were misclassified as two local government systems and one private system.
- A system associated with a veteran's hospital facility in New York is misclassified as a private system.
- A system associated with a private religious facility in New York was misclassified as a local government system.
- Systems in Illinois owned by one of the holding companies are misclassified as local government systems.
- Numerous cooperative (mutually owned) systems are misclassified as local government systems.
- Our own Michigan State University system is misclassified as a local rather than a state governmental system.

Current methods of reporting can blur system owners and operators. The SDWIS data also do not clearly identify system service territories or distinguish between the location of systems and their owner entities. We found 243 system postal addresses that were outside of our Great Lakes region; three systems indicated an address in the state of their incorporation (Delaware), which could obscure ultimate accountability. We also found that ownership data for some systems were out of date and therefore inaccurate as to

Table 2. Summary comparison of reported and recoded Great Lakes states data by SDWIS ownership categories (Fourth Quarter 2019).

SDWIS classification	As reported		As recoded		Net difference	
	N systems	Pop. served (000)	N systems	Pop. served (000)	N systems	Pct.
Local government	6,241	60,199	6,375	62,998	+134	2.1%
State government	106	351	115	394	+11	10.4%
Federal government	34	162	36	162	+2	5.9%
Subtotal	6,381	60,712	6,526	63,554	+147	2.3%
Private	4,831	10,205	4,789	10,205	-42	-0.9%
Public/private	117	1,906	0	0	-117	-100.0%
Native American	82	123	82	123	0	0.0%
Unclassified	0	0	14	2	+14	-
Total systems	11,411	72,945	11,411	72,945	0	0.0%

Source: USEPA (SDWIS, 4Q2019) and authors' analysis.

Table 3. Summary of recoded systems for the Great Lakes states data by SDWIS ownership categories (Fourth Quarter 2019).

		As recoded (moved to)						
		Local govern-ment	State govern-ment	Federal govern-ment	Native Amer.	Private	Public/ private	Total
As recorded (moved from)	Local government	0	2	0	0	164	0	166
	State government	2	0	2	0	2	0	6
	Federal government	1	2	0	0	1	0	4
	Native American	0	0	0	0	0	0	0
	Private	228	11	4	0	0	0	243
	Public/private	70	0	0	0	47	0	117
	Total moves	301	15	6	0	214	0	536

Source: USEPA (SDWIS, 4Q2019) and authors' analysis.

particular type. For example, the operator of record for 13 small well systems in an area of New York had passed away in 2014, and ownership had been transferred to a foreign private holding company (Suez North America) in 2017.

A major takeaway from the coding process is that many of the systems that we classify as ancillary CWSs have characteristics of both community and noncommunity water systems. Ancillary systems account for one-third of the sample and the vast majority of these are nongovernmental. Their reported populations served for the governmental and nongovernmental categories are comparable and together account for less than 2% of the total population served. Depending on research design and controls, ancillary systems

may or may not exert statistical significance in modeling, but they are very relevant from a policy perspective.

For some ancillary systems, reported populations appear disproportionately (even implausibly) high relative to their structure and function.[16] Examples can be found for systems in the military bases, correctional, educational and research, parks and forests, recreational, and commercial types. Some correctional facilities might report total inmate capacity. Some universities might report their student dormitory capacity or actual residents (some of which may reside year-round, as defines a CWS) as well as their total faculty and staff who do not reside on campus. Some parks, commercial enterprises, and recreational facilities (including a casino resort, a racetrack, and steakhouse), might report all patrons, visitors, and timeshare owners in addition to any permanent resident staff. We find, for example, a local park with a population of nearly 10,000 and a state park exceeding 30,000. Yet many ancillary systems may confine their count to residents only, thereby potentially understating population served.

If internal logic checks are lacking at the point of data entry, the current methodology of self-reporting would allow for inaccuracies in SDWIS. We suspect inconsistencies and potentialities in counting and comingling permanent (residential), nontransient, and transient service populations. In a few cases, systems that we classify as ancillary may not actually fit squarely within the statutory definition of a CWS based on population served; that is, they may be misclassified noncommunity systems (with no year-round residents). It is also conceivable that based on their functionalities, some NCWSs could meet the criteria of CWSs.

We do not know whether individual state regulatory policies, including how systems are defined and monitored, might also affect population reporting. Although based on observation and relative statistical impact, we are somewhat less concerned about the primary systems, we cannot know the extent to which they might also be affected. Utility systems only know their active service connections; their population served must be estimated and methods may vary. Verifying system service populations is far beyond the scope of this study, as is devising a corrective strategy to ensure consistency in the database. Our only minor corrective adjustment was to set the population served for all 36 wholesale-only systems to "not applicable."[17]

In sum, in light of our coding experience, we lack confidence in how systems are classified by SDWIS in terms of ownership, even at the highest level (governmental vs. nongovernmental). Based on our review of the anomalies in the database, we also question the validity and reliability of the population data among the ancillary types and their next-level aggregation. Although our recoding provides a more accurate picture of ownership, the self-reported SDWIS data as constituted cannot be used in research to draw firm conclusions regarding system scale across the affected system types, in this analysis or elsewhere.

[16] We dubbed this the "parks and recreation" problem.

[17] Wholesale-only systems appear in SDWIS as serving 0 or 1 customers; we identified and separated 36 wholesale-only systems, of which 18 had reported their population served as "1".

9.0 Key Findings for Water System Types

Having recoded the systems according to our classification scheme, we can examine the distribution of systems by type, considering both the number of systems and reported population served, with the caveat that reported populations for some ancillary systems are likely overstated.

Appendix E provides our detailed census of systems by level and type for the Great Lakes sample. For our sample, as shown in Table 4, governmental and nongovernmental systems account for 57.9% and 42.0% of the systems, and 87.3% and 12.7% of the SDWIS reported population served, respectively. Applying the second level of our typology, a key finding of this research is the striking asymmetry between the presence of *governmental primary* systems (56.0%) and *nongovernmental ancillary* systems (31.9%) in the sample.

The third classification category is applied in Figure 5 to illustrate differences in population served. The specific types with the largest total number of systems in the sample are municipalities (4,300, 37.7%) mobile home developments (2,320, 20.3%), and retail districts and authorities (1,601, 14.0%) followed distantly by rental properties (414, 3.6%), owner associations (402, 3.5%), publicly traded companies (332, 2.9%), and cooperatives and mutual companies (230, 2.0%).

Some interesting findings among the system types are:

- Based on population served, the smallest system in the sample were two mobile home parks (10, which is below the federal threshold for CWSs) and the largest is the New York City system (8.3 million).
- By our SDWIS-based count, the northern Great Lakes region is home to 82 tribal nation systems serving a total population of 123,323.
- We found a total of 36 wholesale-only systems (34 governmental, one for-profit, and one not-for-profit).
- A total of 64 systems in the sample (12.6% of the for-profit systems) are owned by two for-profit foreign equity companies (Corix Utilities, Canada, and Suez North America, France).
- The public trust model is uncommon but very interesting, and Citizens Utilities in Indiana stands out as a prominent example in the region and nationally

Table 4. Distribution of Great Lakes water systems by general types (4Q2019)

	Primary		Ancillary		Unclassified		Total	
	N	% of total	N	% of total	N	%	N	%
Number of systems								
Governmental	6,388	56.0%	220	1.9%			6,610	57.9%
Nongovernmental	1,153	10.1%	3,636	31.9%			4,789	42.0%
Unclassified					14	0.1%	14	0.1%
Total systems	7,541	66.1%	3,858	33.8%	14	0.1%	11,411	100.0%
Reported pop. served (000)*								
Governmental	63,087	86.5%	587	0.8%			63,674	87.3%
Nongovernmental	8,525	11.7%	745	1%			9,270	12.7%
Unclassified					2	0.0%	2	0.0%
Total population	71,612	98.2%	1,273	1.8%	2	0.0%	72,945	100.0%
Purchased water systems	N	% of sys	N	% of sys	N	% of sys	N	% of sys
Governmental	1,920	30.1%	46	20.9%			1,966	29.8%
Nongovernmental	197	17.1%	144	4.0%			341	7.1%
Unclassified					1	7.1%	1	7.1%
Total purchased systems	2,217	20.1%	190	8.2%			2,308	20.2%

Source: USEPA (SDWIS, 4Q2019) and authors' analysis. *Populations served are inconsistently reported in SDWIS; population served for wholesale-only systems was set to n/a (from reported as 0 or 1).

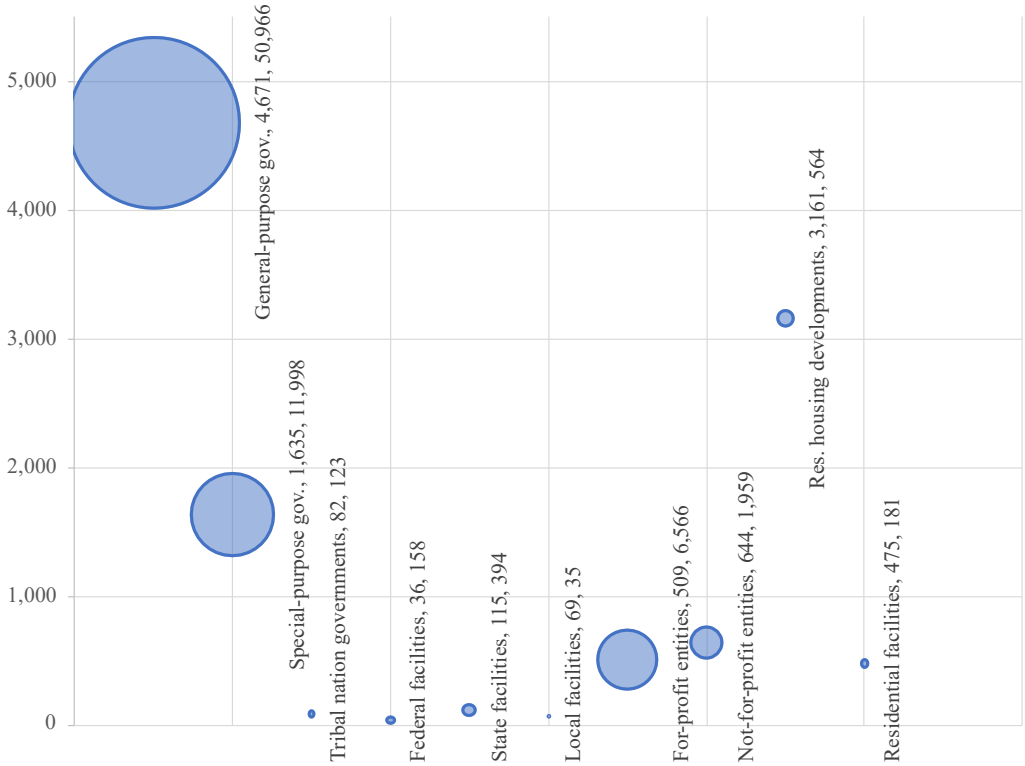


Figure 5. Water systems in the Great Lakes states by categories and population served (level 3).

As summarized in Table 5, The skewness shown confirms a structural feature that is familiar to sector observers and relevant to both research and regulation. In the Great Lakes region, as across the United States, a relatively large number of CWSs serve a small share of the population served and a relatively small number of CWSs serve a large share of the population served. About 80% of the systems are in the smaller two size categories and more than 80% of the population served are in the larger two size categories.

The distributions across the major categories are also as expected. Although the nongovernmental primary systems in the region are about evenly divided between for-profit (4.5%) and not-for-profit systems (5.6%) systems, for-profit systems serve a larger share of the population served (9.0% compared to 2.7%). Across the states, for the primary systems, we also observe the dominance of governmental systems in Michigan, Minnesota, and Wisconsin and a greater presence of non-governmental systems operating on a for-profit basis in Illinois, Indiana, New York, Ohio, and Pennsylvania. This distribution is reflected in the economic regulatory frameworks of these states.

In Table 6, we take a closer look at population served for the governmental and nongovernmental primary systems, which are structured and function as public utilities, to inform our understanding of the sector. The skewness of the distribution of systems is apparent again in the contrast between populations served. The average population served by primary systems is about 9,496; the median is about 500. Overall, it is apparent that governmental primary systems have *system-level* scale advantage over nongovernmental primary systems, serving average (median) populations of 9,929 (1,500) compared to 7,406 (367). Publicly traded and foreign equity companies may have higher averages, but not medians; and those for independent companies are considerably smaller. We note that the lower average population served for the nongovernmental systems is still heavily weighted by Citizens Utilities (Indiana), a public trust.

Larger water systems are generally advantaged because economies of scale are achieved mainly in production, treatment, and operations. Scale can also be captured regionally in the form of interconnection and bilateral wholesale arrangements between geographically proximate systems. A summary for the Great Lakes region is provided in Table 5. Unfortunately, SDWIS does not allow for reporting multiple water sources and their proportions. However, across the Great Lakes states, about 20.2% of systems (28.1% of primary systems) report that they purchased water is their primary source, accounting for about 21.1% of population served. Two states that stand out are Illinois (42.0% of systems and 37.1% of population served and Michigan (20.4% of systems and 65.1% of population served).[18]

[18] The Great Lakes Water Authority (GLWA) in the Detroit metropolitan area serves a population approaching 3.8 million (nearly 40 percent of the state's population) through 88 partners across 8 counties and 112 communities (www.glwater.org).

Table 5. Governmental and nongovernmental water systems in the Great Lakes region by population category.

Number of systems	<500	501 to 3,300	3,301 to 10,000	10,001 to 100,000	>100,000	Total	% of total
Governmental primary	1,556	2,797	1,084	892	59	6,388	56.0%
Governmental ancillary	102	75	34	11	0	222	1.9%
Nongovernmental primary	664	288	88	98	15	1,153	10.1%
Nongovernmental ancillary	3,373	256	4	1	0	3,634	31.8%
Unclassified	14	0	0	0	0	14	0.1%
Total	5,709	3,416	1,210	1,002	74	11,411	100.0%
Percentage of total	50.0%	29.9%	10.6%	8.8%	0.6%	100.0%	
Population served							
Governmental primary	411	4,110	6,456	24,408	27,702	63,087	86.5%
Governmental ancillary	17	118	186	265	0	587	0.8%
Nongovernmental primary	124	400	550	3,187	4,265	8,525	11.7%
Nongovernmental ancillary	439	252	23	32	0	745	1.0%
Unclassified	2	0	0	0	0	2	0.0%
Total	993	4,880	7,215	27,891	31,967	72,945	100.0%
Percentage of total	1.4%	6.7%	9.9%	38.2%	43.8%	100.0%	

By nearly double, we find that governmental primary systems are more likely to purchase water (30.1%) as compared to nongovernmental systems (17.1%); counties (70.1%) and townships (54.1%), lead in purchasing. Even about half of wholesale-only systems purchase their water before reselling it to retailers. We also find that compared to for-profit systems, not-for-profit systems are more likely to purchase water, including about half of the cooperative systems (47.8%). In particular, for-profit publicly traded companies, which dominate the category, are less likely to realize production economies in this manner. This might be explained by the inclination toward private capital investment, on which returns are earned. We emphasize, however, that scale economies cannot be fully assessed without more information about system characteristics, including water production and sales to both the residential and nonresidential classes.

Finally, we also find additional clues about how systems are connected by common ownership (utilities) and/or management (operators). Due to diseconomies of water distribution, physical connection is not always feasible or practical. In SDWIS, about 710 “administrative names” account for a total of 2,330 systems and about 651 “organizational names” for a total of 1,991 systems; even with a majority of these potentially overlapping, we estimate that as much as a quarter of our sample are connected by these means.

Table 6. Primary water system types by population served and purchased water as primary water source.

Primary systems*	Systems		Population*			Purchased water	
	N	%	Pop. reported (000)	Average pop.	Median pop.	% of systems	% of pop.
Governmental primary	6,388	84.7%	63,087	9,929	1,500	30.1%	22.9%
Municipalities	4,300	57.0%	48,245	11,220	1,641	23.5%	19.8%
Townships	279	3.7%	1,859	6,662	1,778	54.1%	79.8%
Counties	92	1.2%	862	9,374	1,496	70.7%	67.7%
Wholesale-only auth.	34	0.5%	na	na	na	47.1%	na
Retail districts and auth.	1,601	21.2%	11,998	7,494	1,100	42.3%	23.4%
Tribal entities	82	1.1%	123	1,504	318	0.0%	0.0%
Nongovernmental primary*	1,153	15.3%	8,525	7,406	367	17.1%	11.5%
For-profit entities	509	6.7%	6,566	12,925	758	12.8%	10.0%
Wholesale-only cos.	1	0.0%	na	na	na	100.0%	na
Publicly traded	332	4.4%	5,653	17,028	1,109	15.1%	7.0%
Foreign equity	64	0.8%	730	11,403	720	12.5%	29.6%
Independent companies	112	1.5%	183	1,631	373	5.4%	27.2%
*Not-for-profit entities**	644	8.5%	1,959	3,046	210	20.5%	16.4%
Wholesale-only NFP	1	0.0%	na	na	na	100.0%	na
Cooperatives	230	3.0%	945	4,089	845	47.8%	31.2%
Owner associations	402	5.3%	139	346	140	4.7%	14.7%
Public trusts*	11	0.1%	875	79,539	700	18.2%	0.6%
Total	7,541	100.0%	71,612	9,496	500	28.1%	21.5%

Source: USEPA (SDWIS, 4Q2019) and authors' analysis. *Populations served are inconsistently reported in SDWIS. The calculation of average and median population served excludes wholesale systems. *Excluding the public trust category, which is heavily weighted by Citizens Utilities, drops the average population served for nongovernmental primary systems to 6,710 (367 median) and for not-for-profit systems to 1,715 (209 median).

Taken together, our research findings make the point that U.S. water systems are capturing production economies of scale as well as (more limited) management economies, more than many researchers, pundits, and policymakers recognize. We note that unless a system is physically absorbed in its entirety, administrative consolidation (through ownership or operation) does not reduce the number of regulated water systems, or avoid system-level reporting, as some presume. Moreover, the optimal number of systems for the country is unknown but likely higher than might be expected give spatial and cost realities across this vast country.

10.0 Conclusions and Recommendations

SDWIS data figure prominently both research and regulation of the U.S. water sector but are mismatched to these purposes. From our experience, however, SDWIS data should not be used "as is." Problems with the accuracy of regulatory compliance data are well known, but the validity and reliability of the data used to identify and differentiate water systems are equally important to the interpretation of research findings and the development of regulatory policy. Our coding experience exposes an ambiguous coding, classification, and reporting system that is inconsistently implemented. Moreover, aggregating and mislabeling highly disparate systems (namely, the "private" category) mask relevant skewness and asymmetry based on structures and functions.

Quantitative researchers are drawn to large and open federal databases and assume their authority and veracity. However, as used in much of the research literature, the available data are not sufficiently valid or reliable to establish the identity and enterprise status of water systems (including ownership, operation, and interconnection) and its influence on performance, including regulatory compliance. Findings from this body of research can oversimplify or even misrepresent both problems and solutions, which in turn can misdirect policy. Utilizing the SDWIS database requires cleaning and correcting for both reported classification and service population issues that permeate the data. Unless they openly address and control for these and other issues, researchers should be far more circumspect about utilizing SDWIS data to draw conclusions or make recommendations from their research, perhaps especially about making assertions about the superiority or inferiority of any type.

Of course, for the long term, the ideal response is not to fix the data but to fix the system by which the data are generated for use, which requires both time and process. For the immediate term, we urge clarity about the structure and function of the water industry in terms of material differences between (1) water utilities and water systems, as well as how systems are already connected via owners, operators, and water sources; (2) between governmental and nongovernmental systems, and in the latter group, between for-profit and not-for-profit systems; and (3) between primary and ancillary systems, and their functional asymmetry. Reducing the ambiguity around these issues will enable more meaningful comparative analysis, with appropriate statistical controls.

Beyond these concerns, we recommend a deeper evaluation of whether SDWIS reporting, and even the broader regulatory framework in which it is embedded, is sufficient for intended purposes. We highlight three issues centering on the available options for reporting data, and thus its accuracy and usefulness to researchers and regulators.[19] First, reported service populations do not reflect the reality that many regulated CWSs are actually hybrid systems serving no transient and transient service populations of various kinds of facilities in addition to their permanent residents; these distinctions matter to regulatory jurisdiction and rules for monitoring, reporting, and operations. Second, CWSs may utilize multiple water sources, but can only report their primary source and not its

[19] The data collection instruments of some states imply interest in these variables, but consistent implementation was not apparent.

proportion; interconnection for supplemental or backup purposes will be underreported. Third, wholesale-only water systems will record a service population of 0 or 1, which technically might not qualify them as CWSs; their impact varies but counting their populations served and that of the retail systems they serve would be double counting. Moreover, may CWSs provide both wholesale and retail services, but the details are not recorded.

The SDWIS rubric is somewhat arcane and well-entrenched, so reform is path-dependent (as the SDWIS Prime experience showed). Full reform that touches, for example, on the criteria by which water systems are defined and regulated, raises the specter of federal legislation. Nonetheless, SDWIS can be improved to establish a clear, transparent, and uniform classification scheme that would collect relevant information but also reduce ambiguities and the chance of self-reporting errors (e.g., by simplified coding and internal logic and validation algorithms). Appendix F provides our wish list of information that might be provided through SDWIS in support of both analysis and policymaking. The list is long, but within the USEPA's statutory purview, but should not overly burdensome.

System reporting should be sufficient for effective regulatory oversight but not overly burdensome, which could undermine compliance. Indeed, a robust reporting system with an emphasis on positive user experience could facilitate monitoring and reporting. However, SDWIS cannot be everything to all researchers, as sometimes imagined. Studies of system economics and governance, for example, may require sampling and other means of data collection. The use of consistent system identifiers across governmental agencies and public policy areas would be helpful for these purposes.

Our proposed classification scheme, informed by our knowledge of the sector and this research, is incremental and parsimonious, and preserves many SDWIS conventions. While we regard our research as a substantial step in the right direction, we also view it as the beginning of a process of adaptation and improvement. Additional refinements could further parse structures and functions perhaps resulting in additional classification levels, but risk reaching diminishing returns in terms of utility. Any classification scheme and typology will benefit from vetting and refinement by researchers, stakeholders, and policymakers in the sector to ensure its robustness, usefulness, and staying power.

For future research, an obvious next step would be to recode and reclassify all of SDWIS. This could inform the refinement of the classification scheme and typology as well as a SDWIS redesign. With a clearer understanding of the industry, future research can focus more clearly on performance drivers and the application of alternative methods and models. Research should aim toward the fundamental purpose of SDWIS, that is, to advance regulatory compliance and the provision of safe drinking water to the people that systems of various types serve.

Acknowledgments

The authors thank the reviewers for their suggestions. This research was supported by the Institute of Public Utilities (ipu.msu.edu) with partial support from the Environmental Science and Policy Program (espp.msu.edu) at Michigan State University.

Bibliography

1. Allaire, M., Wu, H., & Lall, U. (2017). National trends in drinking water quality violations. *Proceedings of the National Academy of Sciences*, *115*(9), 2078-2083.
2. GAO. (2011). Unreliable State Data Limit EPA's Ability to Target Enforcement Priorities and Communicate Water Systems' Performance. *United States Government Accountability Office,* Rep. No. 11-381.
3. Greiner, P. T. (2016). Social Drivers of Water Utility Privatization in the United States: An Examination of the Presence of Variegated Neoliberal Strategies in the Water Utility Sector. *Rural Sociology, 81*(3), 387-406.
4. Grigg, N. S. (2018). Classifying Drinking Water Systems to Improve Their Effectiveness. *Journal-American Water Works Association*, *110*(11), 54-62.
5. Grooms, K. K. (2016). Does water quality improve when a Safe Drinking Water Act violation is issued? A study of the effectiveness of the SDWA in California. *BE Journal of Economic Analysis & Policy*, *16*(1), 1-23.
6. Josset, L., Allaire, M., Hayek, C., Rising, J., Thomas, C., & Lall, U. (2019). The US Water Data Gap – A Survey of State-Level Water Data Platforms to Inform the Development of a National Water Portal. *Earth's Future*, *7*(4), 433-449.
7. Lee, M. Y. A., & Braden, J. B. (2007). *Consolidation as a Regulatory Compliance Strategy: Small Drinking Water Systems and the Safe Drinking Water Act* (No. 9772). American Agricultural Economics Association. Annual Meeting, July 29-August 1, 2007.
8. Marcillo, C. E., & Krometis, L. A. H. (2019). Small towns, big challenges: Does rurality influence Safe Drinking Water Act compliance? *Journal-American Water Works Association Water Science, 1*(1), 1-8.
9. McDonald, Y. J., & Jones, N. E. (2018). Drinking Water Violations and Environmental Justice in the United States, 2011–2015. *American Journal of Public Health*, *108*(10), 1401-1407.
10. McGavisk, E., Roberson, J. A., & Seidel, C. (2013). Using community economics to compare arsenic compliance and noncompliance. *Journal-American Water Works Association*, *105*(3), E115-E126.
11. Montgomery, A. W., Lyon, T. P., & Zhao, D. (2018). Not a Drop to Drink? Drinking Water Quality, System Ownership, and Stakeholder Attention. *Emerald Publishing Limited, Social Movements, Stakeholders and Non-Market Strategy* (pp. 207-245).
12. Oxenford, J. L., & Williams, S. I. (2014). Understanding the causes for water system failure. *Journal-American Water Works Association, 106(*1), E41-E54.
13. Oxenford, J. L., & Barrett, J. M. (2016). Understanding Small Water System Violations and Deficiencies. *Journal-American Water Works Association, 108(*3), 31-37.
14. Pape, A. D., & Seo, M. (2015). Reports of water quality violations induce consumers to buy bottled water. *Agricultural and Resource Economics Review*, *44*(1), 78-93.
15. Pennino, M. J., Compton, J. E., & Leibowitz, S. G. (2017). Trends in drinking water nitrate violations across the United States. *Environmental Science and Technology, 51*(22), 13450-13460.

16. Pierce, G. & Gonzalez S. (2017). Public Drinking Water System Coverage and Its Discontents: The Prevalence and Severity of Water Access Problems in California's Mobile Home Parks. *Environmental Justice, 10*(5), 168-173.
17. Rahman, T., et al. (2010). Determinants of Environmental Noncompliance by Public Water Systems. *Contemporary Economic Policy, 28(*2), 264-274.
18. Rose, J. (2019, April 3). The US Drinking Water Supply is Mostly Safe, but That's Not Good Enough. *The Conversations*. https://theconversation.com/the-us-drinking-water-supply-is-mostly-safe-but-thats-not-good-enough-115028
19. Rubin, S. J. (2013). Structural changes in the water industry from 2001 to 2011. *Journal-American Water Works Association*, *105*(3), E148-E156.
20. Rubin, S. J. (2013). Evaluating Violations of Drinking Water Regulations. *Journal-American Water Works Association, 105*(3), E137-E147.
21. Schaider, L. A., Swetschinski, L., Campbell, C., & Rudel, R. A. (2019). Environmental justice and drinking water quality: are there socioeconomic disparities in nitrate levels in US drinking water? *Environmental Health*, *18*(3), 1-15.
22. Scott, T., Moldogaziev, T., & Greer, R. (2017). Drink What You Can Pay For: Financing Infrastructure in a Fragmented Water System. *Urban Studies,* 1-17.
23. Switzer, D., & Teodoro, M. P. (2017). The Color of Drinking Water: Class, Race, Ethnicity, and Safe Drinking Water Act Compliance. *Journal-American Water Works Association, 109*(9), 40-45.
24. Switzer, D., Teodoro, M. P., & Karasik, S. (2016). The Human Capital Resource Challenge: Recognizing and Overcoming Small Utility Workforce Obstacles. *Journal-American Water Works Association, 108*(8), E416-E424.
25. Teodoro, M. P. (2014). When professionals lead: Executive management, normative isomorphism, and policy implementation. *Journal of Public Administration Research and Theory*, *24*(4), 983-1004.
26. Teodoro, M. P., & Switzer, D. (2016). Drinking from the talent pool: A resource endowment theory of human capital and agency performance. *Public Administration Review, 76*(4), 564-575.
27. Teodoro, M. P., Haider, M., & Switzer, D. (2018). U.S. Environmental Policy Implementation on Tribal Lands: Trust, Neglect, and Justice. *The Policy Studies Journal, 46*(1), 37-59.
28. Tilton, K., & Rosen, J. S. (1997). Reinventing a national drinking water information system. *Journal-American Water Works Association*, *89*(2), 16-114.
29. U.S. Environmental Protection Agency, Office of Inspector General (2017). "EPA Is Taking Steps to Improve State Drinking Water Program Reviews and Public Water Systems Compliance Data." Report No. 17-P-0326.
30. U.S Environmental Protection Agency (1997). Community Water System Survey 1995. *Overview*.
31. U.S Environmental Protection Agency (2002). Community Water System Survey 2000. *Overview*.
32. Wallsten, S., & Kosec, K. (2008). The effects of ownership and benchmark competition: An empirical analysis of US water systems. *International Journal of Industrial Organization, 26*(1), 186-205.

Appendix A: Details of the proposed water system typology

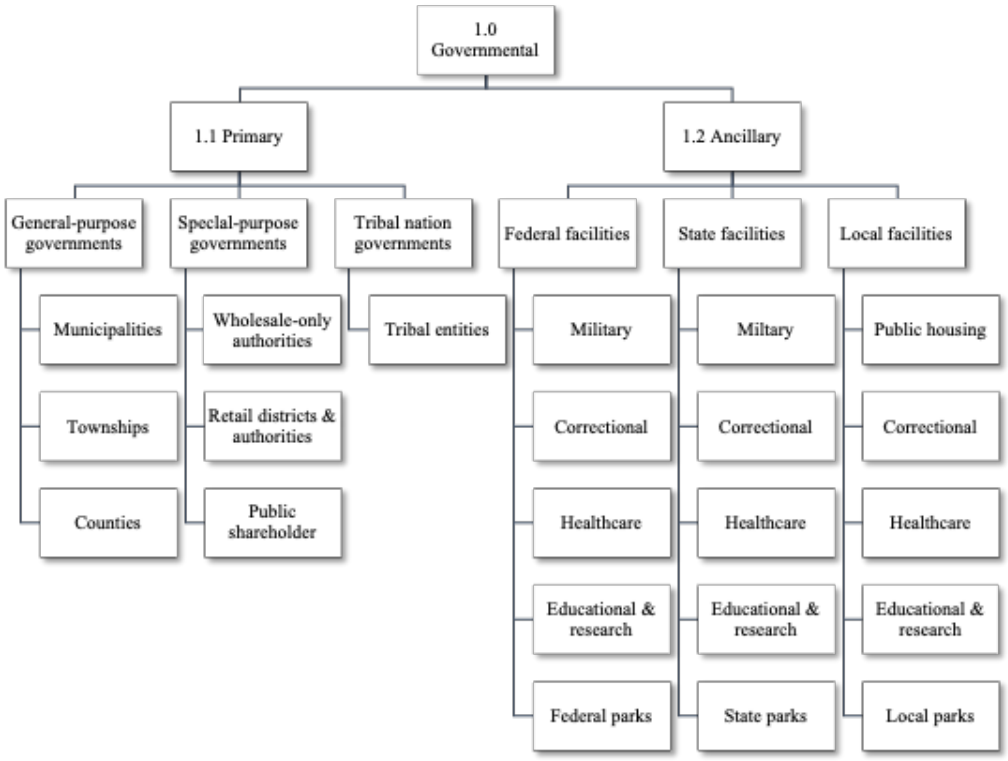


Figure A1. Water system organizational chart: governmental (levels 2, 3, and 4).

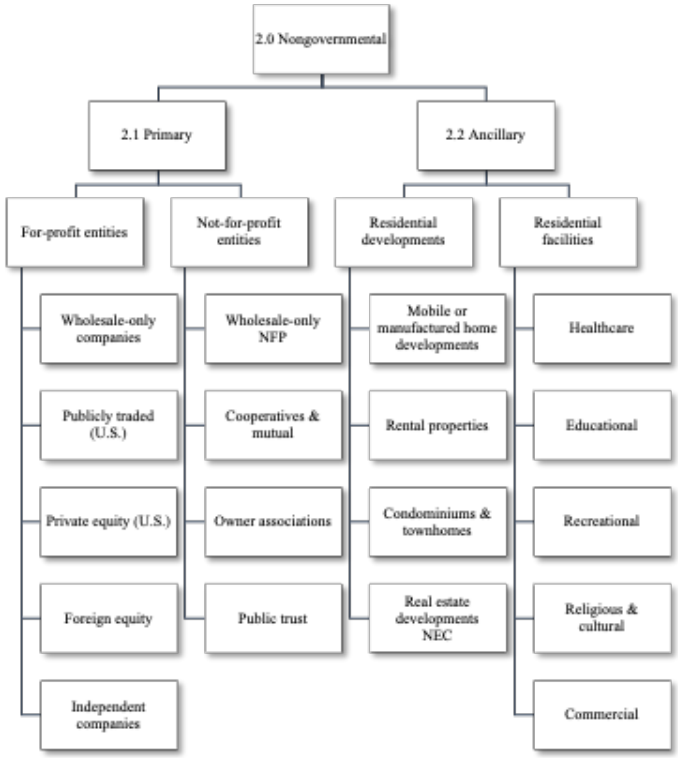


Figure A2. Water system organizational chart: nongovernmental (levels 2, 3, and 4).

Appendix B. Definitions used in the proposed water system typology

1.0	GOVERNMENTAL
1.1	Governmental Primary. Units or divisions of government or public corporations whose primary function is water service (i.e., utilities).
1.1.1	*General-purpose governments.* Authoritative sub-state units of government established to provide general wholesale and/or retail services, including service to proximate systems.
1.1.1.1	Municipalities. Includes cities, towns, and villages. Governmental division or department.
1.1.1.2	Townships. Administrative subdivisions of counties.
1.1.1.3	Counties. Administrative subdivision of states.
1.1.2	*Special-purpose governments.* Authorized sub-state and sub-local units of government established to provide service to a specified area.
1.1.2.1	Wholesale-only authorities. Government-owned water producers (retail population served = 0).
1.1.2.2	Retail districts and authorities. Provide retail service and possibly sales for resale
1.1.2.3	Public shareholder. Wholly owned governmental corporation (formally) with government as sole shareholder.
1.1.3	*Tribal nation governments*
1.1.3.1	Tribal entities. Native American (Indian) owned and operated.
1.2	Governmental Ancillary. Units or divisions of government or public corporations whose primary function is not water service but that provide water to a residential population.
1.2.1	*Federal facilities*
1.2.1.1	Military bases. Includes armed forces bases and academies.
1.2.1.2	Correctional. Includes prisons and detention centers.
1.2.1.3	Healthcare. Includes veteran and psychiatric.
1.2.1.4	Educational and research. Includes federal laboratories and other research facilities.
1.2.1.5	Parks and forests. Includes nature preserves and refuges.
1.2.2	*State facilities*
1.2.2.1	Military bases. Includes national guard bases and academies.
1.2.2.2	Correctional. Includes prisons and detention centers.
1.2.2.3	Healthcare. Includes veteran and psychiatric.
1.2.2.4	Educational and research. Includes universities, laboratories, and other research facilities.
1.2.2.5	Parks and forests. Includes nature preserves and refuges.
1.2.3	*Local facilities*
1.2.3.1	Public housing. Housing authorities.
1.2.3.2	Correctional. Includes jails, correctional centers, and detention centers.
1.2.3.3	Healthcare. Includes county homes.
1.2.3.4	Educational and research. Includes local schools and research facilities.
1.2.3.5	Parks and forests. Includes nature preserves and refuges.
1.2.3.6	Developments. Commercial and industrial under local government management.

2.0	NONGOVERNMENTAL
2.1	Nongovernmental Primary. Private for-profit or not-for-profit entities whose primary function is water service (i.e., utilities).
2.1.1	*For-profit entities.* Investor-owned and operated; typically subject to state economic regulation.
2.1.1.1	Wholesale-only companies. Investor-owned water producers, regardless of the type of equity (retail population served = 0)
2.1.1.2	Publicly traded (U.S.). Entities owned by holding and other companies with equity shares sold on the U.S. stock exchanges; excluding wholesale-only.
2.1.1.3	Private equity (U.S.). Entitles owned by U.S. holding and other companies with investment funds not publicly traded; excluding wholesale-only.
2.1.1.4	Foreign equity. Entitles owned by non-U.S. entities (private or other) with equity shares not publicly traded on the U.S. stock exchanges.
2.1.1.5	Independent companies. Primarily intrastate retail operations with one or more systems and one or more private investors or proprietors; includes LLCs; excludes publicly traded.
2.1.2	*Not-for-profit entities.* Not owned or operated to generate profits.
2.1.2.1	Wholesale-only NFP. Member-owned or mutual water producers (retail population served = 0) (typically, 501(c)(12)).
2.1.2.2	Cooperatives and mutual companies. Member-owned entities that provide retail service and possibly sales for resale (typically, 501(c)(12)).
2.1.2.3	Owner associations. Includes homeowner, condominium, and property associations (typically, 501(c)(4)).
2.1.2.4	Public trusts. Assets held in a charitable public trust for the benefit of the public.
2.2	Nongovernmental Ancillary. Entities whose primary function is not water service but that provide water to a residential population.
2.2.1	*Residential housing developments*
2.2.1.1	Mobile-home developments. Mobile or manufactured home developments (MHPs).
2.2.1.2	Rental properties. Apartments and other rentals, including nongovernmental senior and low-income housing.
2.2.1.3	Condominiums and townhomes. Excludes owner associations.
2.2.1.4	Real estate developments NEC. Includes single-family and mixed density developments not elsewhere classified.
2.2.2	*Residential facilities*
2.2.2.1	Healthcare. Includes hospitals, retirement homes, and other providers of assisted living and supportive care.
2.2.2.2	Educational. Includes non-state colleges, universities, seminaries, and boarding schools.
2.2.2.3	Recreational. Includes residential resorts, timeshares, hotels, camps, and event venues.
2.2.2.4	Religious and cultural. Includes convents, monasteries, and retreats.
2.2.2.5	Commercial. Includes working farms and vineyards.
3.1	UNCLASSIFIED SYSTEMS.

Appendix C. Detailed steps in the system coding process.

1. Working over the course of a year, we downloaded the SDWIS CWS data for the first quarter of 2019 to create a working file for the year; once the data were available, we matched identification numbers to update the database and make it current as of the fourth quarter of 2019 for further research purposes.
2. Our databased included the following SDWIS parameters: system identification number, primacy agency code, system name, administrator name, organization name, address, email address, populations served, counties served, cities served, water source, and owner type.
3. We matched the SDWIS data to primacy agency data for each of Great Lakes states, which we expect to be more detailed and accurate; for our purposes, the augmented data quality appeared higher for Illinois, Indiana, Ohio, Pennsylvania, and Wisconsin.
4. We sorted (and resorted) the data by key terms, synonyms, and abbreviations embedded in PWS names; we also sorted by state and population served.
5. We verified and documented the identity of systems using multiple and hierarchical sources of confirmation; three coders worked both independently and collaboratively to "solve" for classifications of individual systems.
6. Our coding rules prioritized the most authoritative publicly available documentation, such as federal Internal Revenue Service (IRS) tax filings (particularly for not-for-profit systems) and regulatory reports, filings, and tariffs, and governing bylaws.
7. When official documentation was unavailable, we looked for reliable "clues" in the data based on names plus other confirming sources (such as websites or news items); naming is more pragmatic and reliable for some types of systems (such as mobile home parks).
8. We contacted a small number of water system administrators by phone to confirm their identity directly; in some cases, we also contacted state officials to gain perspectives on the systems under their jurisdictions.
9. We applied coder judgment until the research team determined that diminishing returns (or maximum exhaustion) had been reached in terms of the verification process. While accuracy cannot be guaranteed, we are confident that our process resulted in a higher degree of data veracity than is apparent in SDWIS itself or related research.
10. At the conclusion of the coding process, we were left with some cases where the system itself or its parent entity some could not be verified as any type of functioning water system, thus leaving them unclassified.

Appendix D. Great Lakes states compared to SDWIS by population served (Fourth Quarter 2019)

	<500	501 to 3,300	3,301 to 10,000	10,001 to 100,000	>100,000	Total
Number of systems						
Illinois	627	680	224	216	10	1,757
Indiana	270	300	123	81	5	779
Michigan	715	300	155	128	7	1,380
Minnesota	478	308	89	87	3	965
New York	1,422	557	162	133	17	2,297
Ohio	471	382	161	147	12	1,173
Pennsylvania	1,124	484	186	130	16	1,940
Wisconsin	555	309	98	78	4	1,044
Tribal nations	47	21	12	2	0	82
>Total for Great Lakes	5,709	3,416	1,210	1,002	74	11,411
	50.0%	29.9%	10.6%	8.8%	0.7%	100%
>Total for SDWIS	26,946	13,324	5,026	3,911	441	49,648
	54.3%	26.8%	10.1%	7.9%	0.9%	100%
>Difference in percentage	-4.2%	3.1%	0.5%	0.9%	-0.2%	0.0%
Population served (000)*						
Illinois	144	907	1,304	5,699	3,975	12,031
Indiana	44	435	726	2,190	1,550	4,946
Michigan	112	552	948	4,042	1,667	7,321
Minnesota	90	401	492	2,543	933	4,458
New York	230	804	998	3,676	12,523	18,221
Ohio	91	546	963	3,856	4,826	10,282
Pennsylvania	186	761	1,144	3,952	5,401	11,443
Wisconsin	88	439	592	1,909	1,091	4,120
Tribal nations	7	34	57	25	0	123
>Total for Great Lakes	993	4,880	7,215	27,891	31,967	72,945
	1.36%	6.69%	9.89%	38.24%	43.82%	100%
>Total for SDWIS	4,545	19,157	29,573	112,755	144,540	310,568
	1.46%	6.17%	9.52%	36.31%	46.54%	100%
>Difference in percentage	-0.10%	0.52%	0.37%	1.93%	-2.72%	0.0%

Source: USEPA (SDWIS, 4Q2019) and authors' analysis.

Appendix E. Census of Great Lakes water systems by type (4Q2019)

		Systems		Reported pop. served		Purchased water systems	
		N sys.	% of total	N pop.	% of total	N. sys	N. pop.
1.0	GOVERNMENTAL SYSTEMS	6,608	57.9%	63,674	87.3%	1,966	14,583
1.1	Governmental Primary	6,388	56.0%	63,087	86.5%	1,920	14,428
1.1.1	*General-purpose governments*	4,671	40.9%	50,966	69.9%	1,227	11,623
1.1.1.1	Municipalities	4,300	37.7%	48,245	66.1%	1,011	9,555
1.1.1.2	Townships	279	2.4%	1,859	2.5%	151	1,484
1.1.1.3	Counties	92	0.8%	862	1.2%	65	584
1.1.2	*Special-purpose governments*	1,635	14.3%	11,998	16.4%	693	2,805
1.1.2.1	Wholesale-only authorities	34	0.3%	na	na	16	na
1.1.2.2	Retail districts and authorities	1,601	14.0%	11,998	16.4%	677	2,805
1.1.2.3	Public shareholder	0	0.0%	0	0.0%	0	0
1.1.3	*Tribal nation governments*	82	0.7%	123	0.2%	0	0
1.1.3.1	Tribal entities	82	0.7%	123	0.2%	0	0
1.2	Governmental Ancillary	220	1.9%	587	0.8%	46	155
1.2.1	*Federal facilities*	36	0.3%	158	0.2%	12	54
1.2.1.1	Military	15	0.1%	134	0.2%	4	46
1.2.1.2	Correctional	7	0.1%	10	0.0%	2	3
1.2.1.3	Healthcare	9	0.1%	10	0.0%	5	5
1.2.1.4	Educational and research	4	0.0%	5	0.0%	0	0
1.2.1.5	Parks	1	0.0%	0	0.0%	1	0
1.2.2	*State facilities*	115	1.0%	394	0.5%	27	86
1.2.2.1	Military	4	0.0%	6	0.0%	2	5
1.2.2.2	Correctional	62	0.5%	131	0.2%	14	32
1.2.2.3	Healthcare	28	0.2%	20	0.0%	4	1
1.2.2.4	Educational and research	18	0.2%	190	0.3%	7	48
1.2.2.5	Parks	3	0.0%	46	0.1%	0	0
1.2.3	*Local facilities*	69	0.6%	35	0.0%	7	15
1.2.3.1	Public housing	10	0.1%	1	0.0%	0	0
1.2.3.2	Correctional	11	0.1%	5	0.0%	3	3
1.2.3.3	Healthcare	43	0.4%	18	0.0%	3	2
1.2.3.4	Educational and research	3	0.0%	1	0.0%	0	0
1.2.3.5	Parks	2	0.0%	10	0.0%	1	10
1.2.3.6	Developments	0	0.0%	0	0.0%	0	0

Appendix E (continued)

2.0	NONGOVERNMENTAL SYSTEMS	4,789	42.0%	9,269	12.7%	341	1,046
2.1	Nongovernmental Primary	1,153	10.1%	8,525	11.7%	197	980
2.1.1	*For-profit entities*	509	4.5%	6,566	9.0%	65	660
2.1.1.1	Wholesale-only companies	1	0.0%	na	na	1	na
2.1.1.2	Publicly traded (U.S.)	332	2.9%	5,653	7.8%	50	393
2.1.1.3	Private equity (U.S.)	0	0.0%	0	0.0%	0	0
2.1.1.4	Foreign equity	64	0.6%	730	1.0%	8	216
2.1.1.5	Independent companies	112	1.0%	183	0.3%	6	50
2.1.2	*Not-for-profit entities*	644	5.6%	1,959	2.7%	132	321
2.1.2.1	Wholesale-only NFP	1	0.0%	na	na	1	na
2.1.2.2	Cooperatives and mutual companies	230	2.0%	945	1.3%	110	295
2.1.2.3	Owner associations	402	3.5%	139	0.2%	19	21
2.1.2.4	Public trust	11	0.1%	875	1.2%	2	5
2.2	Nongovernmental Ancillary	3,636	31.9%	745	1.0%	144	66
2.2.1	*Residential housing developments*	3,161	27.7%	564	0.8%	126	51
2.2.1.1	Mobile home developments	2,320	20.3%	434	0.6%	112	45
2.2.1.2	Rental properties	414	3.6%	49	0.1%	3	3
2.2.1.3	Condominiums and townhomes	136	1.2%	22	0.0%	0	0
2.2.1.4	Real estate developments NEC	291	2.6%	59	0.1%	11	3
2.2.2	*Residential facilities*	475	4.2%	181	0.2%	18	15
2.2.2.1	Healthcare	288	2.5%	49	0.1%	10	5
2.2.2.2	Educational	56	0.5%	81	0.1%	3	7
2.2.2.3	Recreational	75	0.7%	35	0.0%	4	1
2.2.2.4	Religious and cultural	46	0.4%	14	0.0%	1	1
2.2.2.5	Commercial	10	0.1%	2	0.0%	0	0
3.0	Unclassified	14	0.1%	2	0.0%	1	0
	Total	11,411	100%	72,945	100%	2,308	15,408

Source: USEPA (SDWIS, 4Q2019) and authors' analysis. Percentages shown are based on grand totals for systems and population served. *SDWIS-reported populations served for some systems are likely overstated, particularly for ancillary systems; population served for wholesale systems was set to n/a from reported as 0 or 1.

Appendix F: SDWIS coding wish list

- A PWS identification number that is used consistently across all federal and state administrative agencies and regulatory jurisdictions, including the IRS.
- A PWS name that is current, legally recognized, and consistently used by systems and regulatory agencies (i.e., names associated with the PWS should be updated in a timely manner).
- Legal identity for the owner of record for the system assets with ultimate responsibility for system operations (i.e., governmental, cooperative, association, corporate, or individual owner).
- Ownership of the system by a larger public utility entity, including entities operating in other sectors, such as wastewater or energy.
- Current contact information for the water system (name, physical and email addresses, phone number, and website).
- Tax status and tax identification number for the ultimate owner (i.e., for-profit 1040 vs. not-for-profit 990 or other).
- Responsible operator of record for purposes of SDWA compliance, including current contact information.
- Functional character of water service as either primary or ancillary to another line of business.
- Structural character of water service as wholesale-only, retail-only, or both, specifying proportions.
- Structural character of water service as distribution-only service (no production or treatment).
- Geographic territory (localities) served by the system (i.e., counties and cities) on a retail basis both inside and outside of corporate boundaries.
- Authority of the system entity to raise tax revenues (regardless of use).
- Authority of the system entity issue bonds (regardless of use).
- Identity of an external public or private entity (or partner) contractually engaged to operate the system, if applicable.
- Number of residential service connections and population served
- Number of nonresidential (commercial and industrial) connections.
- Annual volume of water produced and purchased (with identity of wholesaler).
- Annual volume sold to the residential and nonresidential classes.
- Number (or percentage) of residential connections that are metered.
- Frequency of meter reading and billing.
- System’s area in square miles.
- System’s miles of water distribution mains.
- Economic regulation of the system by the state public service commission (PSC).
- System’s aliases with respect to corporate identity (i.e., “doing business as”), as well as name changes (i.e., “formerly known as”).
- System’s nonactive status if applicable (i.e., defunct systems) and form of dissolution.