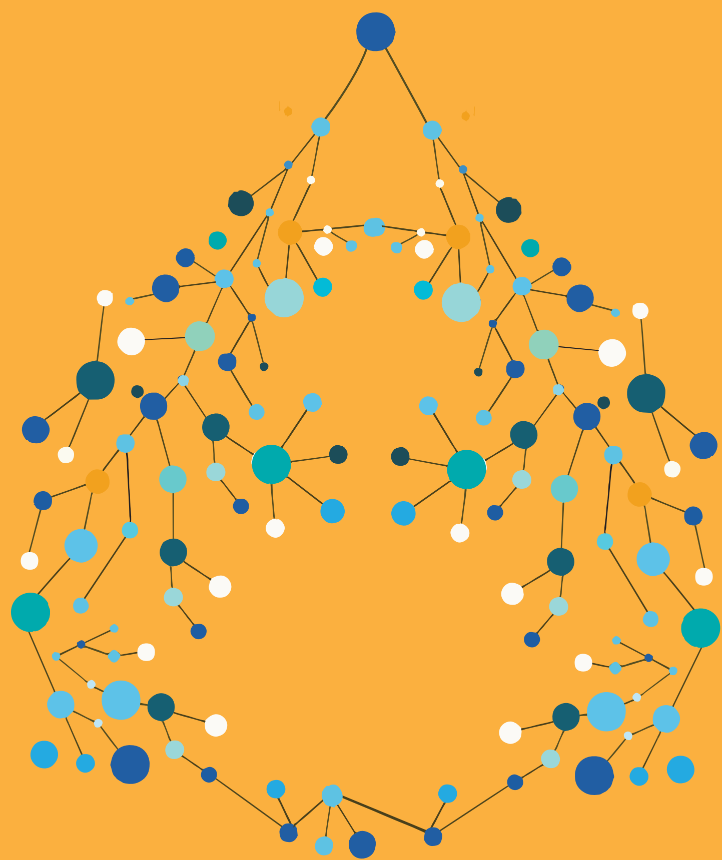
January 2022

Data for Decision-Making

Water and Sanitation in Low-Resource Settings



Contents

Abbreviations and Acronyms

AI	Artificial intelligence
app	Software application, especially as downloaded to a mobile device; may also signify a web-based or desktop computer application
BoQ	Bill of Quantity
CACTUS	Cost and Climate Impacts of Urban Sanitation Technologies
CLARA	Capacity-Linked water supply and sanitation improvement for Africa's peri-urban and Rural Areas
CLTS	Community-Led Total Sanitation
CSO	Civil Society Organization
CWIS	Citywide Inclusive Sanitation
DHS	Demographic and Health Surveys
DMA	District Monitoring Area
Flood-MAR	Flood-Managed Aquifer Recharge
FSM	Fecal sludge management
GIS	Geographic information system
GLAAS	Global Assessment and Analysis of Sanitation and Water
Global South	The Global South is an emerging term, used by the World Bank and other organizations, identifying countries with one side of the underlying global North–South divide, the other side being the countries of the Global North.[1] As such the term does not inherently refer to a geographical south, for example most of the Global South is within the Northern Hemisphere.
GPS	Global Positioning System
GWI	Global Water Intelligence
JMP	UNICEF/WHO Joint Monitoring Programme
LSMS	Living Standards Measurement Study
MDC	Mobile data collection
MICS	Multiple Indicator Cluster Surveys
MoH	Ministry of Health

NRW	Non-revenue water
RRSG	Rethinking Rural Sanitation Guidance
RTU	Remote telemetry unit
SCADA	Supervisory Control and Data Acquisition
SDG	Sustainable Development Goal
SIASAR	Central American Rural Water and Sanitation Information System
UAV	Unmanned aerial vehicle
UN	United Nations
UNICEF	United Nations Children's Fund
USAID	United States Agency for International Development
VANET	Vehicular Ad-Hoc Network
WASH	Water, sanitation, and hygiene
WASHPaLS	Water, Sanitation and Hygiene Partnerships and Learning for Sustainability
WHO	World Health Organization
WPS	Water, Peace and Security partnership

Acknowledgments

This work was generously supported by grants from the Conrad N. Hilton Foundation to The Aquaya Institute. We greatly appreciate the participation of numerous water, sanitation, and hygiene development experts who shared their insights about data science for decision-making, as individually acknowledged in Appendix 2. Alicea Easthope-Frazer (Aquaya Institute) led the effort. Aquaya staff members who contributed to research planning, interviews, data analysis, and reporting included: Clara MacLeod (currently with the World Health Organization), Caroline Delaire, Ranjiv Khush, Edinah Samuel, Jessica Tribbe, Kara Stuart, Katherine Marshall, AJ Karon (currently with McKinsey & Company), Joan Kones (currently with Kenya Water and Sanitation Civil Society Network), Joyce Kisiangani, Chloé Poulin, Karen Setty, and Vanessa Guenther. Andrea Jones (formerly Hilton Foundation) provided valuable input during project inception and dissemination. Contextual perspectives from representatives of several other funding organizations were garnered separately via Aquaya's "Project W" data hub initiative.

Executive Summary

BACKGROUND

Sustainable access to safe and equitable water, sanitation, and hygiene (WASH) is a basic human need that remains unmet in numerous locations. Nations around the world strive to close this gap, at present under the banner of the United Nations Sustainable Development Goal 6. Achieving universal, adequate, accessible, and equitable WASH coverage requires the concerted efforts of professionals from national and local governments, development agencies, civil society organizations, private companies, and research institutions, as well as citizens and community organizers.

A further imperative to advance WASH development emerged during a global crisis, the COVID-19 pandemic. The accompanying global shutdown, which limited traditional, field-based WASH data collection and monitoring, has focused increasing attention on emerging data sources and analytics that could be better leveraged to support WASH improvement efforts. Over the long run, actions to consolidate WASH information resources, reduce one-time use of datasets, and leverage a broader range of data sources (including those previously considered unrelated) will have powerful implications. Accompanying advances in artificial intelligence (AI) analysis methods will also increase capabilities for learning and responding to critical WASH needs.

The goal of this report is to coalesce knowledge about how WASH stakeholders view emerging trends in data science. It represents a planning effort to align data science advances with the most potent WASH needs and demands. Analyzing how various professionals contribute to or could use data science illustrates points of potential engagement that could lead to clearer partnerships and reduce duplicative or ineffective efforts. In cases of severe data paucity, data science activities could be prioritized to offer a baseline for movement toward better-informed decision-making.

APPROACH

More than 65 decision-makers (Appendix 2) were invited to participate in this research, representing a broad cross-section of WASH organizations. Researchers administered semi-structured interview questions during phone or video calls between March and June 2020. The interview guide (Appendix 2) included both general questions for all interviewees and specific questions regarding predetermined data science "use case" hypotheses, tailored as applicable to the decision-makers' professional organizations or roles. Common information needs reported across decision-makers and their organizations were then clustered by topic. Researchers pooled information from multiple interviews as well as related literature to assess and define the characteristics of nine specific data science use cases spanning water, sanitation, communities, programming, finance, and health.

SUMMARY OF USE CASES

Nine use cases were described in depth as pertinent topical examples of using data science to address WASH needs, with complete detail in Annex 1 of this report. These included:

1. Forecasting groundwater quality and quantity— Groundwater supplies are critical to meeting water demand, yet data on their quantity and quality remain hard to come by. Platforms that encourage data access and sharing across political boundaries would help to predict and forestall water supply shortcomings.

2. Reducing non-revenue water (NRW) — Treated water is lost at a high rate in many locations due to both natural and social causes, reducing compensation to water suppliers and straining environmental resources. Addressing this issue through technologies such as remote sensing and telemetry sensors could enhance water service efficiency.

3. Coordinating fecal sludge emptying — Pit latrine and septic tank emptying often takes place ad hoc, leaving pits overflowing, homeowners frustrated, and service providers without work. Coordinating these services using a central application and sensor-equipped vacuum trucks could better align the needs of workers, customers, and regulators.

4. Understanding sanitation costs — Sanitation planning at a city level often introduces excess complexity and ignores the hidden costs of fecal sludge treatment and disposal. Considering the entire sanitation value chain, newer costing applications could use local pricing information to optimize a blend of appropriate options.

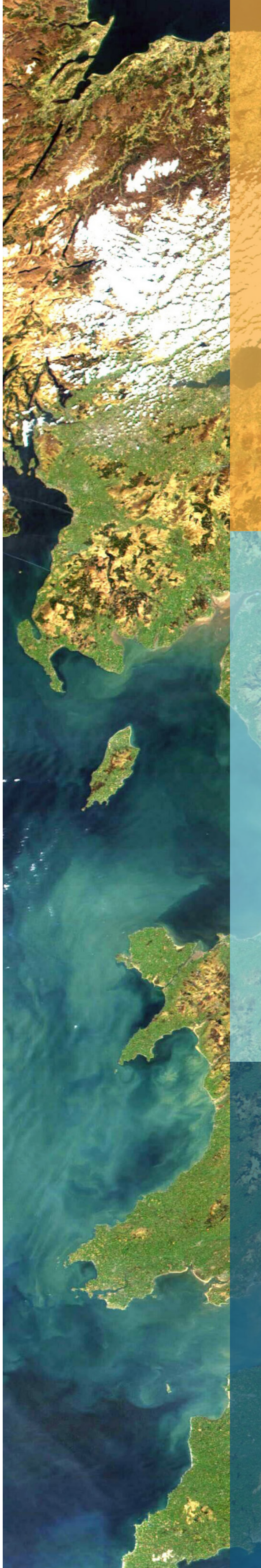
5. Anticipating waterborne disease outbreaks — Retrospective disease surveillance leaves little response time for public health managers to plan or modify prevention and mitigation efforts. Risk mapping and forecasting tools might use algorithms to put decision-makers a step ahead.

6. Interpolating household data — Achievement of global WASH goals relies on household-level access, but descriptive household data are time-consuming to gather and not uniformly available. Advanced data interpolation[1] techniques could be applied to use fewer survey points to generate high-resolution maps and summary statistics.

7. Understanding local contexts through community classification — Tailoring WASH interventions to local community context is both critical to successful programming and notoriously challenging at large scales. Leveraging and combining existing data offers a powerful means to better customize intervention planning.

8. Targeting the poor and vulnerable — Using a single indicator such as annual gross income to qualify household for WASH subsidies may extensively misjudge poverty levels and creditworthiness. Brief, multi-question "smart" surveys offer a pathway to more accurately target financial support using alternative wealth indicators.

9. Evaluating impacts — WASH monitoring and evaluation often falls prey to negative evaluation data at or near the end of projects, when it is too late to respond. Improved, real-time processing of interim or passive data could yield valuable insights to guide investments and clarify success factors.



[1] Estimating values of unknown data points within the range of known data points

RECOMMENDATIONS

Anticipate frequent, albeit often indirect, data use for decision-making. Some professionals disregard the role of data in decision-making simply because it is ancillary or poorly documented. Research shows most data use builds incrementally on existing knowledge and is considered by decision-makers alongside other factors such as social values, costs, and competing interests. In some cases, limited data leads to inaction. Regular engagement between systems and people that produce data and those that use data would ensure its relevance and salience during decision-making windows, whether expected or unexpected.

Normalize sharing to improve the cost-effectiveness of data production. Field collection of primary data may involve extensive startup, implementation, and data processing costs. Standards for data quality assurance and platforms for data sharing are becoming more common, and professional ethics around WASH data sharing are moving toward openness, even in the case of not-quite-successful initiatives. Embracing the power of large datasets for learning about past performance and projecting future performance can aid WASH intervention design at local, national, and international scales.

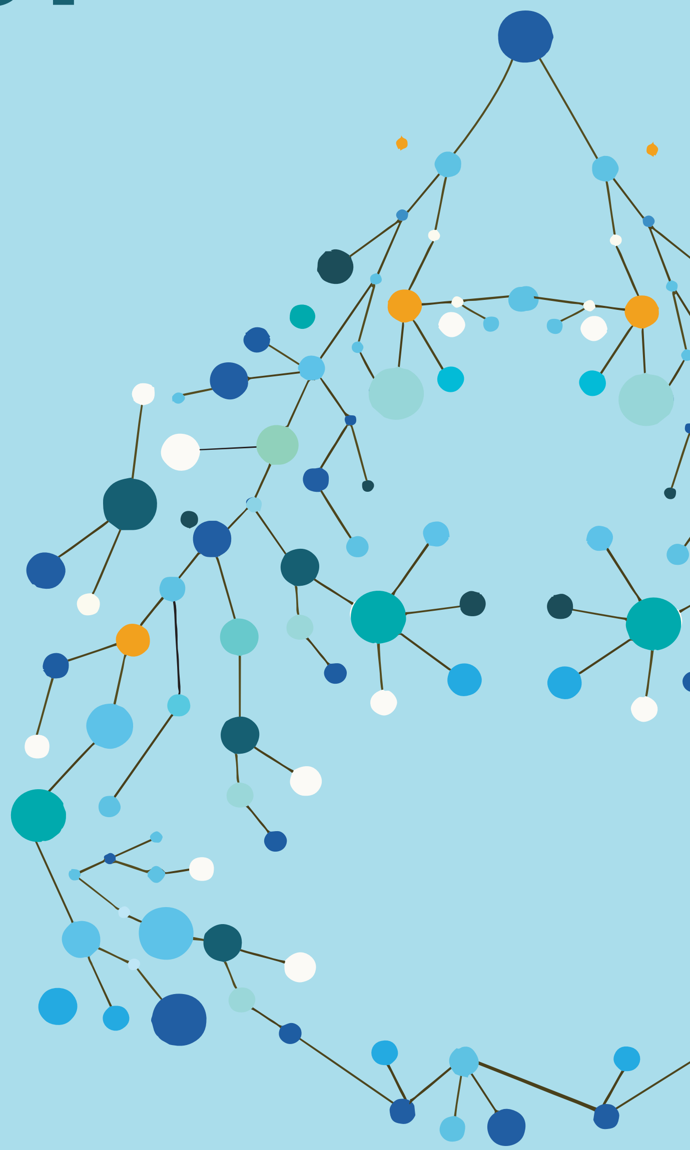
Use advances in automated data recording and analytics to vastly assist, but not replace, human decision-making efforts. Social media, commercial, and satellite data collection platforms have made "big" datasets available, while AI modeling techniques have made it possible to rapidly detect trends that were previously unobservable. Harnessing these tools to support WASH decision-making offers a way to increase the breadth, resolution, and spatial extent of understanding about how individuals and environments interact with WASH services. In turn, building software applications that work in concert with human behavior (e.g., via record-keeping, reminders, or triggers) would ease the time and technical skills otherwise needed to reach conclusions. Still, these products are prone to inherent bias and must be considered through an equity lens within responsible human-guided research and operation practices.

Expand capabilities for reframing the timing of evaluation. Analyzing historical data offers reflective learning value, but limited insight into proactive, forward-looking management. Preventive and adaptive WASH management strategies offer numerous benefits for environmental conservation, cost-efficiency, and public health, as well as improved professional satisfaction and customer service. Moving toward real-time and predictive capabilities would give decision-makers greater lead time to address rapidly emerging issues and pivot when implementation strategies need to be adjusted. For example, algorithms built using historical datasets can be used to communicate real-time WASH service performance and predict failure probability in advance.

Embrace the crucial role of data science in WASH development. At a basic level, WASH development is about providing safe, sustainable, and affordable services to all. Data science approaches for similar development goals (e.g., economy, food, energy, conservation) are already finding their way into the daily lives of WASH professionals and stakeholders. Consciously building on these experiences and advances will ensure WASH development benefits equally from technological progress.

Introduction:

What is data science?



WHAT DOES DATA SCIENCE MEAN?

Data science is a term that encompasses a number of data types and methods for data analysis. Broadly, **DATA SCIENCE** seeks to understand and address problems through the *"useful extraction of knowledge from data via technologies."*[2] The most common approach is identifying predictive patterns through retrospective data analysis. For example, a retailer might project supply and food stocks needed in hurricane-prone areas using historical purchasing patterns prior to past hurricanes.[3]

Data science methods apply to **DIGITAL DATA**, meaning data stored in a digital format (e.g., on a cloud server or hardware), rather than manually recorded (e.g., paper-based). For example, patient medical records are increasingly digitized for access by multiple doctors within the same healthcare system, rather than kept in paper file folders. The concept of **BIG DATA** refers to datasets that are too large for traditional analytical methods. The early 2000s saw the birth and popularization of the term with the 2008 Wired publication, "The End of Theory: The Data Deluge Makes the Scientific Model Obsolete."[4] Using data science methods, big data from multiple sources can be analyzed simultaneously to extract complex insights. For example, advertisements often target individuals or groups of people using their purchase history, location, or interests.[5] Big data are often described with three (or more) "Vs" (Box 1):[6]

- **Volume** – the size of a big dataset depends on the data topic and the user's data management capacity, but can be up to the order of petabytes (quadrillions of bytes);
- **Velocity** – the speed at which data are created and captured, often in real-time; and
- **Variety** – the types of data sources, which might include social media (messages, updates, images), location tracking, sensors, and satellite imagery.[7]

Box 1.

Characteristics of big datasets.

Volume
Amount

Variety
Forms and formats

Velocity
Generation speed

Veracity
Bias, noise, abnormalities.

Validity
Accuracy

Volatility
Time range of retention

Vulnerability
Susceptibility to loss or intrusion

Versioning
Iterations

Value
Variety of uses

Data science has the potential to make data collection more efficient. Modern methods extract information from various technologies that include, but are not limited to: satellite data, location tracking, sensors, and databases (cell phones, health clinics, schools, transportation ridership, etc.). Researchers are increasingly able to collect, analyze, and interpret these data from anywhere in the world in real time.

Data science analytical methods include forms of artificial intelligence (AI) such as machine learning. AI is the concept that artificial objects (i.e., computers) can behave intelligently, meaning they are capable of *"perception, reasoning, learning, communicating and acting in complex systems."*[8] **MACHINE LEARNING** is a type of AI in which machines (i.e., computers) are trained to recognize a pattern in data and communicate a specific response when the pattern is discovered. Over time, the machine "learns" by expanding its knowledge base and improving its performance continuously over time.[9] For example, researchers at Duke University used machine-learning methods to analyze satellite images and identify all solar panels in California cities, lending insight into their total capacity and socioeconomic implications.[10]

WHY IS DATA SCIENCE CRITICAL FOR WATER AND SANITATION DEVELOPMENT?

The health outcomes of poor water, sanitation, and hygiene (WASH) practices are well characterized. According to the 2019 Global Burden of Disease Study, 2.8% of all deaths in 2016 could be attributed to poor WASH conditions.[11] Most high-income countries offer near universal piped water and sewerage in urban areas to safeguard communities from the threat of disease.[12] Low- and middle-income countries, in contrast, have historically lacked the resources to invest in such costly infrastructure everywhere. In harder-to-reach places, governments and civil society organizations (CSOs) have sought to implement low-cost, small-scale solutions. Some interventions demonstrate greater health impacts than others, which may depend in part on contextual factors.[12,13] **Policymakers and investors operating in low-resource settings must therefore carefully consider the trade-offs of introducing any new program, intervention, or product.** As climate change and economic challenges mount, high-income countries also face challenges to maintain updated and accurate information, especially on water quality and quantity.

All governments are now under pressure to meet the specific targets set forth by the United Nations (UN) Sustainable Development Goals (SDGs), particularly Goal 6 targets (Box 2), which aim to ensure access to sanitation and water for all by 2030.[14]

Box 2: Sustainable Development Goal (SDG) 6 Targets

6.1 By 2030, achieve universal and equitable access to safe and affordable drinking water for all.

6.2 By 2030, achieve access to adequate and equitable sanitation and hygiene for all and end open defecation, paying special attention to the needs of women and girls and those in vulnerable situations.

6.3 By 2030, improve water quality by reducing pollution, eliminating dumping and minimizing release of hazardous chemicals and materials, halving the proportion of untreated wastewater and substantially increasing recycling and safe reuse globally.

6.4 By 2030, substantially increase water-use efficiency across all sectors and ensure sustainable withdrawals and supply of freshwater to address water scarcity and substantially reduce the number of people suffering from water scarcity.

6.5 By 2030, implement integrated water resources management at all levels, including through transboundary cooperation as appropriate.
6.6 By 2020, protect and restore water-related ecosystems, including mountains, forests, wetlands, rivers, aquifers and lakes.

6.A By 2030, expand international cooperation and capacity-building support to low- and middle-income countries in water- and sanitation-related activities and programs, including water harvesting, desalination, water efficiency, wastewater treatment, recycling and reuse technologies.

6.B Support and strengthen the participation of local communities in improving water and sanitation management.

Rigorous research, new data sources, and advanced analytical methods will improve the evidence base for choosing the best, most cost-effective solutions to global WASH challenges. In addition to expanding the body of knowledge, research can help address three critical aspects of improving WASH policy and service delivery:
1) program design, 2) monitoring and evaluation, and 3) comparison of cost-effectiveness. Translational research approaches aid the design of specific WASH programs and interventions, development and testing of improved practices, and outcome evaluation.[15]

Nevertheless, there are political, institutional, and behavioral barriers to research and evidence uptake.[16] According to a private sector professional interviewed for this report,[17] *"Fundamentally, resource allocation is political. So, you can have really good data and really good technocrats, but at the end of the day, especially for poor countries that are dependent on projects coming in, there is a lot of pressure to take whatever anyone is offering to demonstrate things are being built and to gain political capital from that. And this utopian ideal of a purely rational decision-making process is too often hijacked by politics, by other power forces and stakeholder dynamics that ultimately decides what happens where."* An urban sanitation researcher[18] noted, *"Political interest hampers [the] urban sanitation decision-making process as well as access to costing data in terms of who should take part in the process…Most of the data users do not understand type of decisions to make from the data."*

Simply because evidence and data exist does not mean decision-makers are able to use the information. **Data needs to be demand driven, accessible, and actionable for decision-makers to use it**.[6] For example, in sub-Saharan Africa, raw water quality data rarely leaves the monitoring agency. Often information reported to regulators and higher levels of government are summary statistics, such as 'percent compliance' and 'total number of tests,' and typically no data are shared with the public.[19] Challenges in reporting data to decision makers include limited data aggregation and analysis and poor enforcement of data-sharing requirements.[19] Decision-makers often lack capacity to analyze and discuss data to support productive decision-making;[20] an interviewee noted, "*Organizations like UNICEF [the United Nations Children's Fund] are very far behind in the understanding of how to use data, so more expertise needs to be brought in to explain the possibilities and improve capacity on interpreting data."*[21]

HOW DOES DATA SCIENCE DIFFER FROM TRADITIONAL COLLECTION METHODS?

Traditional WASH data collection generally requires extensive effort in the field interacting with individuals or the environment. Gathering household-level data requires researchers to hire local enumerators, develop data collection instruments, translate survey instruments, and train teams on how to use the instruments. Enumerators invest substantial labor to identify and survey households, while efficiently developing rapport with each individual to ensure more reliable data. In addition to coordinating field logistics, supervisors may spend time repeating or monitoring survey results to monitor data accuracy and quality. Ideally, local stakeholders and communities are heavily engaged in this process, which benefits the research but adds to the overall time investment.

For environmental monitoring, researchers must acquire and learn how to use the sampling and analysis equipment. Depending on the scope of data collection, large teams might need to be formed and trained as well. **Environmental sampling often requires extensive geographical coverage or a high sampling frequency to sufficiently characterize the research problem.** Blanks, replicates, equipment maintenance, and other quality assurance measures ensure data accuracy but add to the overall time and expense.

Applying data science in conjunction with traditional data collection methods can improve program efficiency and outcomes. Remoted data collection and analysis can be conducted from any location with computing infrastructure and reliable internet access. Smaller teams can manage larger amounts of data to offer insight into WASH systems and interventions. Importantly, data science can improve how data are presented to decision-makers to optimize information availability and certainty.

WHAT DATA SOURCES ARE AVAILABLE?

INTERNATIONAL WATER AND SANITATION MONITORING

International WASH data tracking by the League of Nations Health Organization (predecessor of the World Health Organization [WHO]) began as early as 1930. Monitoring for drinking water access and sanitation coverage increased once the UN established the WHO. In the 1960s, WHO began sending annual questionnaires to ministries of health in participating countries and compiling the results.[22] In 1990, WHO and UNICEF formed the Joint Monitoring Programme for Water Supply and Sanitation (JMP), to centralize production of annual estimated data on global WASH progress. Many indicators from early WHO surveys persist to this day. The late 1990s also yielded the Multiple Indicator Cluster Surveys (MICS) by UNICEF and the Demographic and Health Surveys (DHS) by the United States' Agency for International Development (USAID), resulting in widespread availability of household survey data.[22–24]

Due to quality issues with government-provided data, the JMP began using MICS and DHS household data beginning with their 2000 report.[22] This began shifting the focus of JMP specifically to drinking water and sanitation coverage indicators. MICS was established in 1995, partly to track progress of the World Summit for Children's Declaration and Plan of Action, which called for universal access to drinking water and sanitation by 2000. This was followed by the Millennium Development Goals (MDGs), presented in 2001 but not adopted until 2007, calling for the proportion of people without safe drinking water access and basic sanitation in 1990 to be halved by 2015. JMP became the official progress tracker of the MDGs. As water and sanitation topics gained attention, governments increasingly demanded information on regional policies, challenges, and resource flows. In response, the Global Assessment and Analysis of Sanitation and Water (GLAAS) was born. Implemented by WHO, GLAAS analyzes progress and obstacles within the sector at the national level, primarily using high-quality pre-aggregated WASH data and surveys that assess the nation's enabling environment.[25]

Local water and sanitation monitoring

Water quality monitoring and reporting varies greatly between urban and rural contexts. Urban piped drinking water is typically overseen by municipal utilities, while rural wells fall under the jurisdiction of public health authorities.[26] Water sampling methods are completed either in the laboratory or with field test kits. Field test kits often provide results as a range, or as presence vs. absence, whereas laboratory methods can provide more accurately quantified results and evaluate more compounds.[27] Water suppliers and public health agencies that conduct water quality monitoring typically follow WHO guidelines, national guidelines, and/or applicable regional or internal goals when defining sampling and reporting plans. In low-resource settings, financial constraints often manifest minimal or solely reactive sampling (e.g., following a consumer complaint).[19] When microbial testing kits are unavailable, turbidity or chlorine residual testing is used to mitigate contamination risks.[26] Most water suppliers and health departments digitize data and report to a national regulator; however, a recent study in low-resource settings found that water quality data are seldom used to influence policy decisions, except in response to contamination at the local level or a disease outbreak at the national level.[19] Hygiene practices, wastewater quality, and water quantity are less commonly monitored due to resource constraints, although some data stems from local government oversight (predominantly in urban areas) or research activity.

Emerging technologies for data collection

Meeting SDG 6 will require cost-effective, efficient, and sustainable innovations. This can only be facilitated by *"collection and analysis of increasingly complete and impartial data."*[28] While monitoring coverage has increased dramatically, sizeable data gaps remain in producing reliable sub-national data, particularly in difficult-to-reach areas. **To work around these limitations and hasten data collection, researchers have turned to less hands-on methods**. Emerging methods include mobile data collection, proxy data, machine learning for satellite and aerial imagery, and remote data transmission.

MOBILE DATA COLLECTION

Mobile data collection (MDC) is increasingly common as smartphones become ubiquitous, due to declining prices, increasing network coverage, and better security. A survey during a 2013 MDC event in Paris, France, found that 42% of participants had used MDC, which increased to 75% in 2016 at a similar event in Amman, Jordan [29] MDC can vary in complexity from transmitting photos of broken equipment (e.g., boreholes) to large, structured assessment tools. CartONG, a civil society organization focused on humanitarian uses of geographic data, listed the following benefits of MDC: elimination of inconsistent, impossible, or missing data; reduction of re-keying errors; faster data collection; faster analysis; better quality control; lower costs; and the inclusion of multiple media.[29] Despite advances from traditional paper-based observations, common MDC challenges can include: survey design constraints, hardware failure, connectivity issues, compatibility issues, unfamiliarity of users with the technology, and security and privacy issues.[29] MDC-based innovations used by WASH development stakeholders include mWater,[30] CommCare,[31] ODK Aggregate/ODK Central,[32] Magpi,[33] Akvo Flow,[34] Dharma Platform,[35] Poimapper,[36] and Mobenzi Researcher,[37] among others.[29] A particularly intriguing innovation is Akvo Caddisfly, which connects pocket-sized water-testing devices directly to a smartphone, eliminating data entry errors.[28]

USEFUL PROXIES IN EXISTING DATA

Data for some indicators can be difficult to collect. For example, accurate poverty data are challenging to obtain at scale. Households themselves may not even maintain these records. **To accurately estimate poverty and other hard-to-determine variables, researchers have begun searching for proxies within existing datasets.** One method, the Proxy Means Test, weights information on multiple household characteristics (typically more than a dozen) to estimate poverty.[38,39] For example, variables may include building material of the home, livestock ownership, parental education level, number of children, or car ownership. Researchers have also realized promising results using single variable proxies. Poverty levels have been estimated using mobile phone data usage, for instance, by observing the association between data use and income levels.[40,41] An emerging effort seeks to build a single poverty prediction model using multiple information sources, such as satellite, environmental, mobile, and census data.[42]

SATELLITE REMOTE SENSING

The spatial data captured in satellite imagery can be used to produce high-resolution maps of various indicators at a large scale. For example, recent studies have used satellite imagery to estimate infrastructure decline and public service availability in slums; quantify water quality parameters (e.g., chlorophyll, cyanobacterial pigments, suspended matter) in water bodies; classify land use; monitor rainfall, temperature, and crop production; and identify population groups that may be vulnerable to food insecurity.[28] Specific to WASH, the World Bank recently used remote sensing in Nigeria to compare household survey data with land use classifications from satellite imagery to estimate water and sanitation indicators, such as service access.[28] Nighttime light data from satellite imagery has been used to estimate economic activity, since areas with artificial lights from buildings or transportation infrastructure tend to be wealthier.[43–46] Daytime satellite imagery has been applied to estimate household wealth using roofing material.[47,48]

Unmanned aerial vehicles (UAVs, or drones) are another emerging data collection and mapping technology. UAVs are particularly useful for deriving high-resolution 3D information of surfaces such as watersheds.[28] They have also been used to monitor natural resources, such as biomass, forests, and vegetation. **This cost-effective approach substantially increases temporal and spatial resolution compared to traditional monitoring methods.**[28]

REMOTE MONITORING AND COMMUNICATION TECHNOLOGIES

Technologies such as "smart" water meters, water pump sensors, and latrine motion detectors can expand the scope and reliability of WASH data collection.[28] High quality water meters can reduce water losses, improve service consistency, and increase consumer trust. Water pump sensors can shorten the time interval between breakdown and repair, as shown in Rwanda by the company SweetSense.[28,49] Accelerometers and motion sensors can be used to monitor household water consumption, community adoption of latrines, and the use of handwashing stations. The Passive Latrine Use Monitor, also commercialized by SweetSense, detects warm-bodied movement in a latrine stall to monitor use. To study and promote handwashing, SmartSoap (developed by Unilever) is a bar of soap with an embedded accelerometer that detects use.[28] Finally, Mercy Corps used motion-detector-based latrine sensors and water-flow sensors to measure the prevalence of handwashing after latrine use.

Integrated and automated data collection technologies improve service delivery, offer transparency, and reinforce accountability from water providers and governments.[28] For example, Supervisory Control and Data Acquisition (SCADA) systems arose around the 1960s, as different industries sought to efficiently and automatically control their equipment, particularly over long distances. SCADA systems used in present day water treatment plants evolved in the 1980s and 1990s, made possible with Local Area Networking (LAN) technology. LAN allowed systems from different vendors to communicate, opening the door to have many different devices connected to a single network. **These systems allow water treatment plant operators all over the world to gather and analyze data in real time.** Sensors and controllers interface directly with the water treatment plant and distribution machinery, while workers monitor a communications control center, using graphical user interfaces for high-level supervision of day-to-day processes.

HOW HAVE OTHERS DEVELOPED WASH DATA FOR DECISION-MAKING?

The UNICEF/WHO JMP reports are the official data sources of UN-Water and used for decision-making across the UN.[50] **The JMP itself does not act on survey results, but the data are used widely by other stakeholders.** Sanitation and Water for All (SWA), a partnership of governments, CSOs and other stakeholders, relies on JMP and GLAAS findings for their mutual accountability mechanism. International development organizations, such as WHO, UNICEF, World Bank, and CSOs, can use JMP data for multi-country comparisons. JMP data are used in the United Nations Development Programme's Human Development Report, the World Water Assessment Programme, the Ibrahim Index of African Governance and UN Habitat's slum population analyses. WHO country offices may follow up with national governments about JMP findings, although national governments are usually interested in more granular, sub-national data within their country.

Beyond simply sharing data, some policy-makers proactively incorporate data into their decision-making processes. Two states in the United States, California and Texas, have state government-led initiatives to leverage data for water management and identify needs across their state. For example, California's Water Boards developed a regulatory framework that tracks water quality data from regulated users and automatically enacts enforcement in response to noncompliance. In addition to government-led initiatives, private companies and CSOs are engaged in turning data into actionable information. **Data science presents a business opportunity for the private sector, as well as opportunities to leverage data for advocacy efforts**. Several prominent examples of water-related data science initiatives at global, multinational, national, and state scales appear in Appendix 1. Data sources highlighted within the use cases for this report appear in Table 1.

TABLE 1: Select WASH-related data sources featured in the data for decision-making use cases. For a complete catalog of more than 3000 WASH data sources, see Aquaya's Project W (under development).[51]

NAME	SOURCE	DESCRIPTION	LIMITATIONS
Joint Monitoring Programme (JMP)	WHO/UNICEF	National WASH survey data	Reporting variability, missing data
Multiple Indicator Cluster Surveys (MICS)	UNICEF	Internationally comparable household survey data, includes some household water quality	Not georeferenced
Demographic and Health Surveys (DHS)	USAID	Internationally comparable, nationally representative household survey data	Not available for every country
Global Analysis and Assessment of Sanitation and Drinking-Water (GLAAS)	UN-Water/WHO	National WASH financing survey data	Reporting variability, missing data
WorldPop	WorldPop	Population distribution, demographics, and dynamics in resource-poor settings	Simplified assumptions
Census data	National governments	Population demographics	Availability/questions vary by country
Citywide Inclusive Sanitation (CWIS)	World Bank	Sanitation costing data	Locally specific
Google Maps/Earth	Google	Global surface map data, including topography/elevation	Quality varies by region
OpenStreetMap (OSM)	Crowdsourced	Global surface map data	Less accurate in rural areas
Digital Globe	DigitalGlobe	Satellite and aerial imagery	Requires purchase
AtlasAI	AtlasAI	Satellite and aerial imagery	Requires purchase
British Geological Service (BGS)	UK government	Water table height	Limited free access
International Soil Reference and Information Centre (ISRIC)	ISRIC	Soil type	Fewer observations in some regions
Data for Good	Facebook	Social media data	User bias
Global Enteric Multicenter Study (GEMS)	University of Maryland	Sanitation and hygiene conditions, diarrheal disease data	Select regions (Asia and Africa)
Global Infectious Diseases and Epidemiology Online Network (GIDEON)	GIDEON	Infectious disease data	Requires purchase
Humanitarian Data Exchange (HDX)	Crowdsourced	Data sharing, checking, and visualization	May be context specific

WHY ARE EXISTING DATA APPROACHES FALLING SHORT?

Data shortcomings are plentiful, and vary from stakeholder to stakeholder. Many data shortcomings can be attributed ultimately to a lack of financial resources. Data paucity at the household or consumer level stems from the intensive staff and time requirements needed to collect it. Gathering up-to-date data in humanitarian settings (disaster or conflict-prone areas) is notoriously difficult due to high personnel costs coupled with travel difficulty, unpredictability, and quickly evolving contexts. Routine water quality monitoring is challenging for water utilities in sub-Saharan Africa largely due to a lack of testing materials and partly due to limited regulatory enforcement.[52] Researchers may misinterpret needs and spend time and resources answering irrelevant questions. For example, one expert on water loss noted, *"what is needed is applied research. We don't need yet another algorithm trying to predict leak locations!"*[53] Further, gathering data to demonstrate health impacts requires long-term comparative studies, which are slow, expensive, and challenging (e.g., to retain the same participants). Even when plentiful data exist, their quantity, quality, temporality, and relevance often interfere use in decision-making.

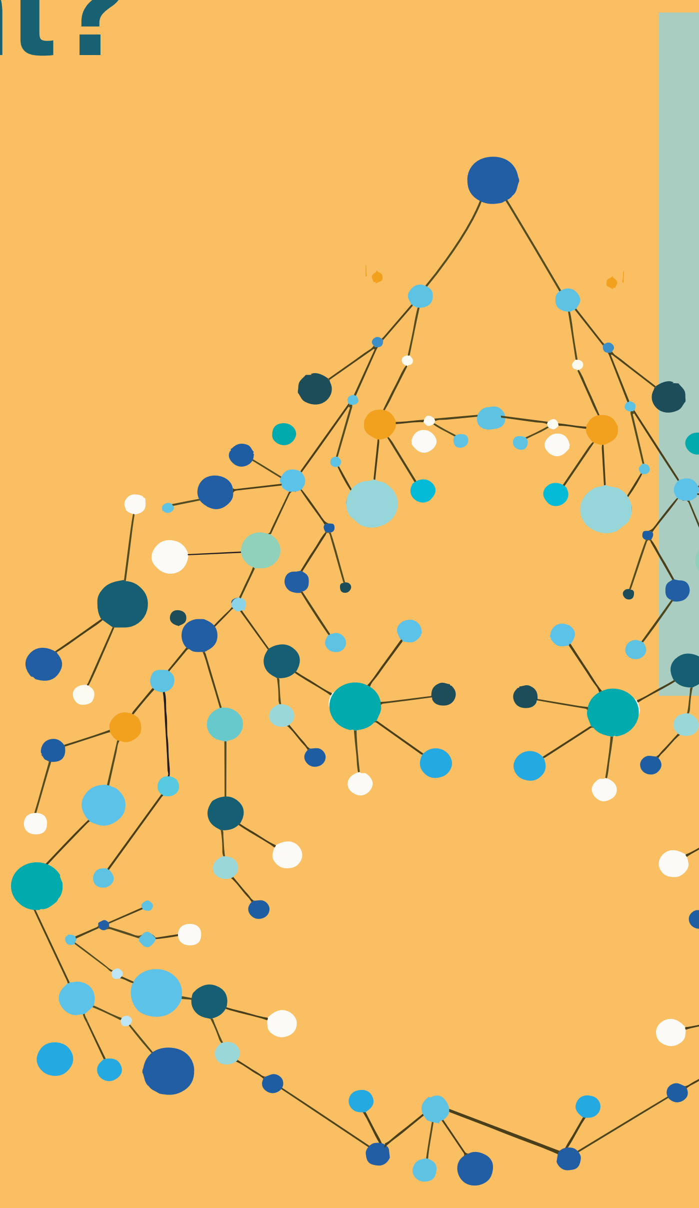
Substantial differences often exist between aggregated national data and household surveys, which vary greatly across contexts. JMP reports rely on raw household survey data and censuses, which are aggregated at a lower resolution (e.g., province or regional level). In an estimated 50% of countries, these are unavailable; thus, JMP is regularly left with "survey reports" or "census reports" prepared by national statistical offices. Further, JMP coverage estimates are calculated using linear regression models, which do not necessarily best fit the data or match methodologies used in-country. Issues also arise with differences in defining the types of facilities that count toward WASH coverage goals. Improved indicators and analytical methods will likely lead to more accurate estimates of coverage and service quality. They will allow for improved tracking of the progress of reaching unserved or underserved populations, and the rate at which populations move up the sanitation ladder. Future reports could include data from regulators, utilities, CSOs, and potentially even individual users.

A wealth of data exists, but it is often collected by different actors and in varying formats. International WASH monitoring and reporting data such as the JMP, DHS, and national censuses usually offer downloadable datasets to corroborate their reports. Household-level data, where available, is obfuscated to protect privacy of respondents. Data collected using emerging technologies are often proprietary and not publicly available, unless included in a report or publication. An exception may be if an academic or other researcher conducts a study on a technology and releases data along with the report or peer-reviewed article. Data collected by utilities, such as water quality data, are typically kept private, although some may produce summary reports for governments and customers. In low-resource settings, utilities are shifting toward digitizing monitoring and evaluation and compliance data, which is not yet universal. Where digitizing does occur, the data are not often subsequently analyzed or used for trend monitoring.

Compiling existing data from multiple sources would make it easier for data to be shared and accessed, and increase collaboration among WASH organizations. A number of data platforms under development can serve as a resource for finding and sharing datasets and convening stakeholders through networking opportunities. These require additional resources to ensure compatibility and validity of data, manage the data and applications, and support users. While these platforms support the organization and categorization of multiple datasets, they may not offer analytical services or aggregate data for "big data" or sector-wide analyses.

Findings:

How could data science aid WASH development?



WHO MAKES WASH DECISIONS?

For this report, we divided decision makers into five primary categories: for-profit or private sector actors, non-profit actors, investors (funders), government actors, and academic actors (Table 2). Each category plays a different role within international WASH development and has different priorities and motivations, as described below. Some organizations may identify with multiple or more specific labels.

FOR-PROFIT ACTORS

For-profit or private sector actors include WASH service and product suppliers, technology companies, consulting companies, and data science startups. **These organizations use data to identify innovation opportunities and offer WASH products and services that are profitable.** For-profit actors use data to advocate for investments in their company and to market their products to customers to maximize profitability. For example, Sanergy, a private Kenyan-based sanitation provider, might use data to better understand customer needs, such as sanitation alternatives and costs, willingness to pay, times of toilet usage, and number of toilet users per household.[54]

NON-PROFIT ACTORS

Non-profit actors include international and local development organizations such as WASH and environmental CSOs (also called NGOs), research or advocacy groups, not-for-profit consulting organizations, professional associations, and religious or cultural institutions. **These organizations use data to implement WASH programs that maximize health and other benefits.** CSOs receiving donor funding use evaluation data to document funding impact and justify new proposals. Additionally, implementers such as World Vision rely on data regarding emerging technologies to ensure projects are using evidence-based approaches. Such groups are often interested in understanding the value of various technologies per dollar invested to aid their decision-making.[55]

DONORS AND INVESTORS

Donors and investors include philanthropic foundations, bilateral and multilateral aid agencies, and social impact investing firms. T**hese organizations use data to determine where the greatest need for investment exists and identify the most cost-effective solutions**. Investors finance pro-poor projects and provide capital to those working to improve WASH outcomes.

GOVERNMENT ACTORS

Government actors include government-operated WASH service providers, ministries, public agencies, and national and local regulators. **These organizations use data to improve WASH service delivery, achieve WASH development goals, develop policies and standards, and manage resources.** Governments that rely on donors may lack the data needed to quickly assess whether a donor-proposed plan responds to the country's priorities.[56]

ACADEMIC ACTORS

Academic actors include universities, students, and individual experts or researchers. **These actors use data to rigorously answer research questions and promote evidence-based solutions that support effective WASH decision-making.**

TABLE 2: Types of decision-makers with examples

FOR-PROFIT

SUBCATEGORY	EXAMPLE ORGANIZATIONS	ASSUMED PRIORITIES
Suppliers	• Private water companies (tanker trucks, sachet/bottled water) • SilAfrica/Kentainers	• Maximizing profits • Innovating within the sector • Creating sustained impact (improved water and sanitation) • Exploring and targeting emerging markets • Demonstrating social responsibility
WASH Technology Companies	• Utilis • XiO • Loowat • Sanergy	
Consulting Companies	• TetraTech • DAI	
Data Science for Development Startups	• AtlasAI • Akvo • Gather	
Big Technology Companies	• Google • IBM	

NON-PROFIT

SUBCATEGORY	EXAMPLE ORGANIZATIONS	ASSUMED PRIORITIES
International Development Organizations	• RTI International • IRC	• Supporting government actors by filling in gaps and sustaining service delivery models that lack capacity • Supporting local economies (e.g.,micro-loans, entrepreneurship)
Associations	• Africa Water Association (AfWA) • International Water Management Institute (IWMI) • Sanitation Learning Hub	
WASH/Environmental CSOs	• Water & Sanitation for the Urban Poor (WSUP) • Charity: Water • Sustainable Organic Integrated Livelihoods (SOIL) Haiti	

INVESTORS

SUBCATEGORY	EXAMPLE ORGANIZATIONS	ASSUMED PRIORITIES
Philanthropic Foundations	• Bill & Melinda Gates Foundation • Mulago Foundation	• Maximizing investment by creating sustained impact • Social or capital returns on investment • Supporting local economies
Social Impact Investing Firms	• Global Innovation Fund • Omidyar Network	
Multilateral/Bilateral Development Organizations	• Department for International Development (DFID) UK • African Development Bank (AfDB)	
Individual Investors		

GOVERNMENT AGENCIES

SUBCATEGORY	EXAMPLE ORGANIZATIONS	ASSUMED PRIORITIES
Service Providers (Utilities)	• National Water and Sewerage Corporation (NWSC), Uganda • National Office for Water and Sanitation (ONEA), Burkina Faso	• Ensuring water and sanitation safety for the public • Creating sustained impact (improved water and sanitation) • Supporting local economies (e.g.,subsidies, entrepreneurship)
Ministries	• Ministry of Health • Ministry of Water and Environment	
Public Agencies	• California Governor's Office of Planning and Research (OPR) • California Water Data Challenge	
Local or National Regulators	• Environmental Protection Agency (EPA) • Water Services Regulatory Board (WASREB), Kenya	

ACADEMIC

SUBCATEGORY	EXAMPLE ORGANIZATIONS	ASSUMED PRIORITIES
Universities	• The Water Institute at the University of North Carolina, Chapel Hill • Swiss Federal Institute of Aquatic Science and Technology (Eawag)	• Understanding which interventions have sustained impact • Identifying where the greatest needs exist • Providing the right people with the right information for decision-making
Students	• Master's degree candidates • PhD degree candidates	
Individual Experts		

WHAT DO DECISION-MAKERS NEED?

Theory around the use of evidence for decision-making in WASH is actively growing and borrows from related fields.[57] Key concepts emphasize the importance of evidence characteristics (e.g., credibility, salience, legitimacy)[6] as well as directly linking supply to demand-driven needs or questions.[58,59] Paynter[58] explored the pitfalls of failing to establish ties between data production and use. These include the addition of "noise," which can increase perceptions of conflict and detract from decision-makers' confidence in evidence.59 Models of evidence use are increasingly multidirectional and rarely capture all real-world nuance, such as indirect uses of information and the influence of heuristics and values in decision-making.[16,57] The Osprey Foundation and Aguaconsult aim to consolidate a framework relevant to WASH, which could be adaptable to understanding varied data use scenarios.

To understand the needs of the sector, we conducted phone interviews with more than 65 individuals representing nearly all subcategories of WASH stakeholders (Appendix 2). Of the 10-person research team, interviewees were matched with one or two interviewers with whom they were professionally familiar. Interviews were semi-structured, with ample room to explore decision-making challenges within their respective roles (full interview guide found in Appendix 3). **The goal of the conversations was to identify the data needs for these key stakeholders to make better decisions.** Prior to the conversations, Aquaya staff members developed hypotheses of which data needs were of greatest priorities across the sector. These conversations confirmed a number of the hypotheses to be true.

Qualitative results from the interviews were aggregated to prioritize areas where data science could add the greatest value to the sector. These priority areas were developed into the use cases summarized below. Common characteristics of decision-makers across the use cases are summarized in Figure 1.

FIGURE 1: Needs of a typical decision-maker

WHICH WASH DATA SCIENCE NEEDS AND SOLUTIONS WERE IDENTIFIED?

Nine data-for-decision-making use cases were developed, drawing from both participant interviews and related literature (Table 3). **All use cases were included in the findings because they (a) are expected to have a high impact on beneficiaries and the environment, (b) have clearly documented demand, and (c) are not restricted to one community or geography.** Detailed methods, objectives, descriptions, decision status quos (present decision-making approaches), demands, other data applications (beyond the use case), existing and upcoming innovations, participants, workflows, data sources, and barriers are available in an accompanying document, *Annex 1: Analysis of Results by Use Cases* to this report, *Data for Decision-Making: Water and Sanitation in Low-Resource Settings*. The findings and select examples for each of the use cases are summarized in this section.

TABLE 3: Overview of use case characteristics

USE CASES	OBJECTIVE	TOPICS	PRIMARY DECISION-MAKERS
1. Forecasting groundwater quality and quantity	Provide user-friendly models and maps that predict groundwater quality and quantity	Water	Governments, civil society organizations, investors, agricultural coalitions
2. Reducing non-revenue water	Provide real-time data on network flow and pressure to reduce non-revenue water	Water	Service providers (water utilities)
3. Coordinating fecal sludge emptying	Provide an application that tracks pit-emptying jobs, locations, and routes	Sanitation	Governments, service providers, customers
4. Understanding sanitation costs	Provide a cost-comparison tool for sanitation interventions	Sanitation	Local governments, service providers, investors
5. Anticipating waterborne disease outbreaks	Provide trajectories of waterborne disease outbreaks to guide targeted prevention and mitigation efforts	Health	Governments, civil society organizations
6. Understanding local contexts through community classification	Provide granular, geo-referenced data on community classification to tailor interventions	Programming	Governments, civil society organizations, investors
7. Interpolating household data	Provide improved, comprehensive household-level data by interpolating actual data	Programming	Governments, civil society organizations
8. Targeting the poor and vulnerable	Provide high-resolution household poverty and creditworthiness information to target WASH subsidies and loans	Programming	Governments, service providers, civil society organizations, investors
9. Evaluating impacts	Provide alternative methods to determine WASH program impacts	Programming	Governments, civil society organizations, investors

ANALYSIS OF RESULTS: USE CASES

1. Water: Forecasting groundwater quality and quantity

KEY INFORMANTS

- Chris Cormency, Chief of WASH, **UNICEF Mozambique**
- Ramon Brentfüherer (Project Manager - Policy Advice Groundwater) and Vincent Post (Hydrogeologist), **German Federal Institute for Geosciences and Natural Resources (BGR)**
- Aude-Sophie Rodella (Senior Economist), Esha Zaveri (Economist), Jason Russ (Economist), and François Bertone (Senior Water Resource Management Specialist, Groundwater), **World Bank, Global Water Practice**
- Neno Kukuric (Director), and Claudia Ruiz Vargas (Researcher), **International Groundwater Resources Assessment Centre (IGRAC)**

BACKGROUND

Groundwater (pumped from aquifers beneath the surface) is an important source of fresh water globally, especially in arid regions or where surface waters are contaminated. One global development goal (SDG 6.1) aims to provide safe, consistently available, and affordable drinking water for all people by 2030, which relies in part on optimizing groundwater supplies. Innovators continue driving forward technological capacity for groundwater quantity and quality forecasting using advanced modeling and machine learning techniques with many practical applications.[60–63] As one example, the Swiss Federal Institute of Aquatic Science and Technology (Eawag) created a global-scale prediction map for elevated concentrations of naturally occurring arsenic in groundwater.[64] It employed a random forest machine-learning model with inputs of 11 geospatial environmental parameters and more than 50,000 field-measured groundwater data points.

Those responsible for ensuring that groundwater quantity and quality adequately meet public needs include national and local governments (often supported by donor and UN agencies), public or private water suppliers, and development organizations. General applications of groundwater data might include:

- Predicting water conflicts that could have severe negative repercussions.
- Designing effective environmental protection regulations and management strategies (e.g., withdrawal limits) to ensure adequate supplies.
- Identifying where projected population and climate shifts will diminish water supplies,[65,66] and how water shortages will conversely displace populations.[67,68]
- Prioritizing investments in geographic areas with mismatched groundwater supply and demand.
- Determining which groundwater supplies are becoming too degraded (e.g., due to salinization[66]) to be used for human consumption or irrigation, or might benefit from recharge or remediation efforts.

- A multilateral agency wants to assess the relative resilience or vulnerability of groundwater supplies to climate change.[21]
- A government agency wants to develop contingency plans (e.g., in case of drought or flooding), or strategize emergency response and recovery efforts.
- A water supply utility wants to plan for infrastructure expansions and maintenance, anticipating potential future water shortages.
- An agricultural cooperative wants to fairly allocate water supplies to ensure long-term sustainability and profitability.
- A CSO wants to collate aquifer assessments across political boundaries and prioritize areas facing the most critical water shortages.

NEEDS

Despite advances in programming, groundwater management challenges remain. In 2015, 31% of the global population (i.e., 2.2 billion people) still did not have access to safely managed drinking water services; the majority of these 2.2 billion people were in Africa.[69] The interviewed decision-makers were most concerned with reaching 100% water supply coverage in rural and low-income urban areas, where infrastructure and data resolution might be limited. When hydrogeology (occurrence, distribution, and movement of groundwater) is not well characterized, for instance, wells that are costly to site and install may dry up over time. Groundwater data paucity limits proactive assessment of such risks; one interviewee[70] noted,

> “Compared to surface water, there is not much data on groundwater. Data on [groundwater] quality is even more scarce than on quantity.”

Even where data abound, organizations may not have the staff capacity to manage and analyze it. Third-party data are often underused, hard to interpret, or limited to developed country locations.[62] Another interviewee[71] stated,

> “The more we can get technology and ICTs [information and communications technology] embedded in monitoring systems, the better. But ensuring effective adoption, use, and maintenance is a challenge.”

Incongruent spatial scales likewise challenge decision makers. Groundwater quality and quantity data are often captured at a localized scale, making it difficult to understand the regional or national status.[72] For example, in the Rift Valley in East Africa, limited data inform the extent and severity of fluoride contamination.[71,73] Field sampling requires extensive time and resources, resulting in often proprietary or small amounts of data with limited application. In contrast, national-level summary data may be inadequate for addressing local issues.

SOLUTIONS

The goal[70] of offering neutral, accessible, and reliable insights about groundwater monitoring data requires partnerships to pool resources, harmonize data collection and analysis methods, develop both off-the-shelf and customizable applications, and control data quality. The following **strategies** and solution examples address barriers by meeting aforementioned **USER REQUIREMENTS**:

Marginalized communities could be included in national or urban planning efforts. The India Water Tool (developed by the World Business Council for Sustainable Development)

(cont.)

offers a **USER-FRIENDLY, OPEN ACCESS** platform compiling more than 20 different data sources to characterize water risks across the country.[74] Stakeholders can access **RAW DATA AND MAPS** related to water stress, groundwater levels, and groundwater quality.

User-friendly online platforms could offer simplified and up-to-date information to proactively evaluate trends and guide cost-effective investments. Global Water Intelligence's WaterData platform offers members a paid service that displays data on the water market by industry or country, **LEVERAGING MULTIPLE DATA SOURCES** and **SAVING THE TIME AND EXPENSE** needed by clients to conduct independent reconnaissance.[75] Aquaya's Project W, under development, will comprise a **CENTRALIZED** hub for virtually all data relevant to WASH (e.g., public health, demographics, environment, service provision, economics, governance), dedicated support for data searches, customized data processing and visualization services, and advanced data **ANALYTICS**.[51]

Organized partnerships could harmonize data collection systems and mapping or analysis applications across political boundaries. The WPS Partnership's machine-learning model for water conflict **FORECASTS UP TO A YEAR IN ADVANCE**.[76] After testing more than 80 indicators, they found variables such as population demographics, crop prices, and seasonal and interannual water variability were among the most relevant indicators. **EXTENDING BROADLY** over Africa, South and Southeast Asia, and the Middle East, the publicly accessible Global Early Warning Tool allows users to overlay areas with a high risk of conflict, below-average precipitation (which recharges aquifers), and other geographic data **CUSTOM-ADDED** by the user. The interactive web-based Global Groundwater Information System (GGIS) from IGRAC aims to provide information on groundwater globally, including well data.[77] Users can upload and assign licensing to **SHARED** datasets.

2. Water:
Reducing non-revenue water

KEY INFORMANTS

- Roland Liemberger, Independent Consultant **specializing in water efficiency**
- Habab Taifour, Senior Water Resources Specialist, **World Bank Ethiopia**
- Philip Oyamo, Senior Project Manager, **Water & Sanitation for the Urban Poor (WSUP) Kenya, an international CSO**
- Isaac Kega (Monitoring and Evaluation Specialist) and Stella Warue (Programme Officer), **Water Sector Trust Fund**, a grant-providing Kenyan State Corporation

BACKGROUND

A major factor affecting water suppliers globally is the difference between the amount of water distributed and the amount actually billed to consumers, called "non-revenue" water (NRW).[78] NRW has two main components: 1) physical losses (e.g., leakage from pipes and storage tank overflows) and 2) commercial losses (e.g., from under-registration of customer meters, data handling errors, or illicit connections). A third source of NRW is authorized but unbilled consumption, which includes water allocated for utility operations, firefighting, or provision to groups unable to pay.[79] Water suppliers in urban areas typically read meters connected to their distribution systems on an intermittent schedule; in some cases they only identify pipe bursts when staff or consumers report them; and in most cases, smaller losses not visible aboveground go undetected.

Water suppliers (utility operators) are primarily responsible for minimizing NRW. Within a utility, managers or technical specialists hold the most expertise in data analytics and interpreting maps and planning drawings, while field personnel with construction and customer service experience form "leak detection teams." Other entities may externally support NRW reduction efforts; these include private companies, national and local governments (often supported by donors and multilateral agencies), and professional associations or partnerships.

Global interest in optimizing NRW levels has grown in recent years. In 2016, the World Bank and International Water Association (IWA) established a global partnership that promotes NRW reductions through performance-based contracts, leading to sizeable investments in countries such as Kenya, Ethiopia, and Vietnam.[80–84] Advances in technology can facilitate NRW optimization efforts. For instance, Utilis' Hydro-Scan technology uses satellite data to detect soil moisture up to 10 feet belowground, combined with an algorithm to characterize it as drinking water from distribution networks.85 Working with a water supplier in Bangkok, Thailand, this approach found more than 2,000 suspected water leaks; on-the-ground verification confirmed 90% of them.[86]

Improved access to high-resolution data on flow and pressure within water distribution networks enables water suppliers to make better informed decisions in the following areas:

- Quantifying baseline conditions and tracking trends in NRW to assess system performance.[53]
- Identifying leakages, pipe bursts, pumping issues, illegal connections, and meter disruptions.
- Prioritizing the most cost-effective remedial actions, such as pipe or meter replacement, pressure management, or connection removal.[87]
- Using past consumption data to anticipate peaks in demand, adapt water production, and avoid service interruptions.[88]
- Improving operational efficiency and conserving water resources.
- Adjusting tariffs to accurately account for NRW and achieve financial sustainability.[83,89]

NEEDS

Poor understanding of the sources, magnitude, and costs of NRW stymies global efforts to address this issue. In some countries, NRW constitutes as much as 35% of produced water.[90] The time and money utilities spend on treating and pumping water that may never reach consumers is not easily recouped. Utilities in the Global South collectively lose approximately 45 million cubic meters of water per day, valued at over 3 billion USD per year.90 Baseline levels of NRW can be established through an audit; however, an interviewee[53] noted,

> “A lot of utilities are lacking a fundamental understanding of their own systems, so it’s hard to build an improvement strategy. […] Basic performance data for utilities is critical to figure out how to improve performance.”

Without regular flow and pressure measurements in several strategic locations throughout the network, utilities cannot perform basic tasks, such as calculating a water balance and estimating physical losses in different sub-sections of the network (often called “district metered areas”) or detecting changes in “minimum night flow” (a standard technique for identifying physical losses).

Institutional barriers, such as weak regulatory enforcement and operational inefficiency, have also proven challenging. Finding the most cost-effective solution for water service providers relies on buy-in from senior leaders, who often do not recognize NRW as a significant issue. An NRW expert[53] noted,

> “the primary bottleneck for NRW reduction in LMICs [low- and middle-income countries] is to ‘change people’s mind’ and get them to ‘get serious’ [about the issue].”

It can further be difficult for utilities to obtain funding to reduce NRW, since there is a misperception of financial risk with minimal impact. Employee and public buy-in is likewise lacking; one interviewee[91] reported,

> “…estimated data could easily be the result of meter readers colluding with households so that the households don’t pay for water consumption. […] There are numerous causes of NRW but this data is not available.”

SOLUTIONS

Providing decision makers with meaningful data on NRW will require strategic shifts in operations and data management approaches. The following **strategies** and solution examples address barriers by meeting aforementioned

Increase the speed as well as spatial and temporal resolution of distribution system data delivery. Emerging technologies from XiO and others apply remote telemetry units (RTUs, electronic devices that transmit data and/or alter the physical state of connected machinery using messages to/from the central control system) to supply **REAL-TIME** data (e.g., pressure readings every few minutes) for water network monitoring.[92–94] Furthermore, cloud-based Supervisory Control and Data Acquisition (SCADA) allows users to both **REMOTELY** monitor and control the condition of field-based equipment across large service areas.[92] In-pipe tools such as the Sahara Leak Detection Platform and augmented reality visualization platforms such as vGIS can help to narrow down the source of water leaks in the field.[95,96]

Analyze large datasets to enhance water suppliers' predictive capacities. AI that use **AUTOMATED** computing can be applied to pipe monitoring and **EARLY DETECTION** of water losses. By detecting spatial and temporal patterns and anomalies in flow and pressure values at different points in the network, algorithms can identify and even classify the cause of water losses (e.g., illegal connections, leaks, pipe bursts, malfunctioning sensors, or abnormal consumption patterns).[97] Water consumption data from "smart meters" (which transmit data remotely) can help to **FORECAST** demand peaks.[98]

Use alert, mapping, or reporting applications to understand progress and provide direction to field teams. Daily or higher resolution RTU data processing with applications such as EcoStruxure™ Geo SCADA Expert,[99] CGI Mosaic,[100] AVEVA OSIsoft,[101] or Metasphere Palette[102] offer a basis for **CENTRALIZING** and standardizing routine and emergency decision-making about utility operations. For periodic use, Utilis' Hydro-Scan can use satellite data produced every two weeks to generate **USER-FRIENDLY** GIS reports.[85] Leakmited provides a similar service combining user-provided maps with historic data and satellite imagery to identify locations vulnerable to leaks.[103]

3. Sanitation: Coordinating fecal sludge emptying

KEY INFORMANTS

- Elizabeth Tilley, Professor, **Malawi Polytechnic University**
- Zaituni Kanenje, Pro-Poor Manager, **NAWASSCO**, Kenya
- Rick Johnston, Technical Officer, **JMP/WHO**
- Richard Cheruiyot, Inspectorate Services Manager, and Brenda Anzagi, Information Communications and Technology Manager, **WASREB**
- Joseph Githinji, General Manager of Operations and Senior Manager of Customer Support, **Sanergy,** Kenya
- Nienke Andriessen and BJ Ward, Eawag, **Fecal Sludge Management Group**

BACKGROUND

As governments strive to achieve SDG 6.2, data on fecal sludge emptying practices and services could improve service provision, public health, and environmental quality. The WHO/UNICEF JMP's definition of safely managed sanitation specifies that fecal waste should either be transported through a sewerage system for off-site treatment or temporarily stored in a safe onsite containment structure prior to transport and offsite treatment.[104] Building or extending sewer networks in rapidly growing urban areas is often costly and depends on favorable terrain, leaving manual or mechanical (e.g., vacuum truck) fecal sludge removal from pit latrines and septic tanks as a primary alternative.

Those responsible for fecal sludge management include national and local regulators,[105] public and private service providers, and local residents, businesses, and other institutions. External support for improved fecal sludge management may come from development organizations and their funders. Typical stakeholders that require fecal sludge emptying and management data include:

- Local government agencies seeking to monitor regulatory compliance of service providers.[106]
- Service providers marketing and implementing their services, who need to know when to visit customers, where to find pit latrines or septic tanks, how to avoid traffic, and where to dispose of fecal sludge.[54]
- Customers requiring emptying services who wish to compare service providers, easily request and pay for services, view the truck's location, and provide feedback.
- Funders deciding how to prioritize investments in fecal sludge management (e.g., extending existing sewer networks, serving difficult-to-access areas).
- Researchers and innovators designing effective treatment procedures.[107]

NEEDS

Formal, regulated fecal sludge management (FSM) services primarily target middle- and high-income households that can afford market prices.[108] Examples from Kenya and Bangladesh highlight the unsafe fecal sludge emptying challenges common in low-income urban areas. In Kisumu, Kenya, approximately 60% of the population uses unimproved sanitation.109 Only about 20% of residents are connected to sewers and 5% to septic tanks.[110] The majority rely on informal manual pit emptying, whereby fecal sludge is dumped directly into the environment or buried onsite. In Bangladesh, nearly half of the 55 million urban residents lack access to sanitation facilities that enable fecal waste to be safely collected and removed for treatment. As a result, huge quantities of fecal waste are dumped into drains and rivers, contaminating the environment and posing serious public health risks.[111]

Waste emptiers are rarely organized into formal associations, customers have limited methods to get in touch with them, and service providers take little initiative to contact previous customers.[105,112] Potential customers interviewed under a sanitation development program for the city of Port Harcourt, Nigeria, noted it was difficult to arrange emptying services,

> “[Vacuum truck operators] can sometimes be difficult to contact, and it takes roughly 48 hours to reach them.”[113]

Meanwhile, vacuum truck operators struggled to obtain enough work:

> “The business does not provide frequent-enough jobs to form an association. We usually only get a job once in every two weeks at the most.”[114]

Even formal associations often lack data on their sanitation emptying business trends. For instance, an interviewee reported[115]

> “The association doesn’t know when the business is in the peak or when it’s low, how much they should charge a household based on distance or amount of solid waste in the pit.”

Further, regulators are unable to keep track of all fecal sludge dumping events. Another interviewee[116] noted,

> “If trucks and dumping sites were geo-localized so that they could be tracked on a dashboard, it would help fight wild dumping.”

SOLUTIONS

Advancing safe management of fecal sludge in dense, urban environments requires improved coordination and information sharing. The following **strategies** and solution examples address barriers by meeting aforementioned **USER REQUIREMENTS:**

Continue building on innovative sector coordination approaches. Water & Sanitation for the Urban Poor (WSUP) developed a mobile app, Pula, offering customer data and truck **LOCATION TRACKING**.[117] Initial piloting in Lusaka, Zambia, and Maputo, Mozambique, however, revealed it was too complex for most users. The Global System for Mobile Communications (GSMA) Mobile for Development (M4D) Utilities Innovation Fund supported development of mobile apps in urban areas in Uganda and Bangladesh that **CONNECT ENTREPRENEURS WITH CUSTOMERS** and coordinate and track service delivery across the sanitation value chain.[118] They also collect payments. Using a “switchboard” model[105,119] in Dakar, Senegal, the national sanitation agency (ONAS) developed a program to support

(cont.)

low-income households' access to mechanized sludge emptying services, train and certify private operators, and run a call center. The call center gathers service requests by phone and requests and shares bids by SMS (**TEXT MESSAGE**), which increased access and lowered prices by 20%.[119]

Leverage existing data sources to enhance sanitation efficiency and safety. In Mubi, Nigeria, researchers applied exclusively **SECONDARY DATA** (remote sensing and GIS) to design a new sewer network.[120] In Nonthaburi, Thailand, researchers logged GPS data from fecal sludge collection trucks to develop algorithms that **OPTIMIZED** pickup and disposal routes. Traveling distances were reduced by half using the improved routing advice.[121] The Greater Warangal Municipal Corporation in India requires emptiers to keep **DETAILED RECORDS** of every job and fit trucks with GPS trackers.[105] To improve garbage collection efficiency, researchers in India simulated optimal trash collection routes using wireless sensor networks (WSNs) on residential dustbins and vehicular ad-hoc networks (VANETs).[122] **SENSING** when bins were full enough for collection and when trucks were filled helped to optimize pickup routes and prevent

Develop a new multi-user application to optimize urban fecal sludge emptying activities. Novel software applications could serve multiple functions depending on the user **INTERFACE**: review service provider

(cont.)

ratings and availability, request pit/tank emptying, list open jobs, identify pit locations due for emptying soon (for targeted advertising), suggest transportation routes and disposal facilities, and report on historical records such as emptying frequency. An online web application synced to the mobile app can serve as a **DATA DASHBOARD** offering downloadable reports. The app should be developed and piloted in a dense urban area with some low-income housing, and **MARKETED** among end users in coordination with a collaborating municipality.

4. Sanitation:
Understanding sanitation costs

KEY INFORMANTS

- Fiona Zakaria, Research Fellow, **University of Leeds**, UK
- Joseph Githinji, General Manager, Operations and Senior Manager, Customer Support, **Sanergy,** Kenya
- Andrea Jones, Independent Consultant

BACKGROUND

A central challenge to meeting the SDG sanitation target is selecting the most appropriate waste management option(s) for each local context.[123] Eawag's Compendium of Sanitation Systems and Technologies describes numerous sanitation solutions, such as conventional sewers, simplified ("condominial") sewers, pit latrines, septic tanks, composting toilets, vault toilets, and container-based sanitation.[124] High-resolution contextual and pricing data could be useful for customizing sanitation services at the neighborhood level, thereby enhancing suitability, cost-effectiveness, and sustainability. To improve data availability for decision-making, innovators have piloted life-cycle costing apps,[125,126] some of which compare sanitation options at a citywide level.[127,128]

Those responsible for selecting appropriate sanitation options include sub-national governments, public and private service providers, donors, implementing organizations (e.g., CSOs), and sanitation researchers. Specific applications of cost and suitability data might include:

- A local government agency needs to compare site-specific life-cycle costs for a mix of sanitation options in a given area over a set time span.
- A public utility needs to run cost-benefit scenarios for upgrading their sewer networks.
- A service provider needs to define sanitation tariffs and subsidy amounts.
- A donor needs to understand how much funding would be required to achieve sustainable sanitation access (e.g., to meet development goals) and compare service delivery options.
- A CSO needs to designate zones within a city where sanitation access is lowest.
- A private company needs to understand to what extent household income and local preferences favor certain sanitation options.[54]

NEEDS

The SDG call for universal access to sanitation by 2030 is not yet on track to be met.[69] Providing for or improving sanitation infrastructure and services, particularly on a city-wide level, is very costly. The lack of accurate cost data, as well as guidance on the characteristics, benefits, and limitations of multiple sanitation solutions, inhibits informed, data-driven choices.[123,129] Hutton suggested,

> "We need to move public sector decision-making more towards the supermarket model, and increase the availability of key information so that decisions can be more rational, consistent, and transparent."[130]

Low quality or limited cost data forces assumptions, leading to "incomplete and potentially misleading results."[131] Decision-makers are challenged to gather comparable data on different sanitation service options, in part due to hesitancy among private providers to share cost information. Trémolet et al.[132] noted that,

> "despite decades of field experience, reliable estimates for the hardware and software costs of sanitation access are still scarce."

Decision-makers typically select sanitation interventions without considering cost-benefit accounting for the full sanitation chain, leaving them prone to failure.[18,130] They often rely solely on the lowest dollar amount needed on a city-wide level (total capital investment) or historical political choices.[130] Manga et al. found very few urban local decision-makers calculate the real costs of managing onsite sanitation systems, which include emptying sludge, transporting it for treatment, and operating treatment facilities.[133] An interviewee[18] further noted a disregard for local preferences,

> "Most initiatives driven by development partners or stakeholders fail because they don't understand the beneficiaries' needs and the processes are not people-centered…"

Most studies of sanitation economics lack the costing data necessary to cover life-cycle costs over the full sanitation chain from facility construction to waste disposal or reuse.[123,129] Typical study limitations consider only one component of the sanitation system or only the first year of operation.[134] When available, sanitation costing data may be very specific to one project or context, material and labor costs may be aggregated, and corresponding metadata (e.g., design drawings and photos) may be missing.[135] Ulrich et al.[135] stated,

> "Estimating capital and operational infrastructure costs is not easy, especially considering all the context-specific and variable factors that determine the total costs."

Other costing needs include distinguishing urban vs. rural areas and converting data to comparable metrics (e.g., cost per capita, unit cost).[123] Ideally, widespread design data for key sanitation options would be compiled into standardized bills of quantities (BoQs; lists of required materials and services).

SOLUTIONS

To enable sanitation intervention comparison and selection for a given urban area, decision-makers require multiyear data processing applications that account for local contexts and pricing. The following **strategies** and solution examples address barriers by meeting aforementioned **USER REQUIREMENTS**:

» **Encourage sanitation life-cycle cost comparison.** In Ethiopia, the CLARA Simplified Planning Tool (SPT) calculates and **COMPARES** full costs (infrastructure, operation, and maintenance) of sanitation systems during early planning phases.[136] Standard design options can be selected, with costs calculated from BoQs. This approach has been adapted to local

(cont.)

contexts in Kenya, South Africa, and Ethiopia. The World Bank's Economics of Sanitation Initiative (ESI) Toolkit uses algorithms to calculate and compare results of various projected solutions across different time periods and areas.[137] Its cost-benefit and cost-effectiveness modules produce **STANDARD INDICATORS** demonstrating socioeconomic return on sanitation investment, and results are easily shared among users. The Community-Led Urban Environmental Sanitation (CLUES) toolbox is an automated, BoQ-based, Microsoft Excel tool that estimates construction and maintenance costs of select sanitation technologies; however, it does not estimate total life-cycle costs and costs of design alternatives may vary widely.[135,138] The Citywide Inclusive Sanitation (CWIS) Services Assessment and Planning (SAP) tool developed by Athena Infonomics with funding from the Gates Foundation systematically considers a range of scenarios and evaluates **TRADEOFFS** relative to equity, sustainability, and safety.[139] It is being tested in urban areas of six countries. Finally, the Climate and Costs in Urban Sanitation (CACTUS) tool produces costs **ACROSS THE SANITATION LIFE CYCLE** (from collection to treatment).[140] It outputs total annualized cost per household and per capita, normalized to a single currency and date.

Leverage local sanitation cost data to inform development progress. The World Bank and UNICEF developed a WASH SDG Costing Tool that calculates the financing gap needed to fulfill the SDG sanitation target at a country level.[141]

(cont.)

The Excel model **CONSIDERS EXPECTED POPULATION CHANGES** between 2015 and 2030 (or a custom baseline year), although its general value assumptions introduce some uncertainty. The International Water and Sanitation Center (IRC) WASHCost life-cycle costing tools for water and sanitation services collect user data on infrastructure components and costs and upload them to the mWater app.[30] **AUTOMATED** analyses describe the gap between existing services and full coverage at the desired service level, as well as the affordability and adequacy of tariffs paid by households.[142]

Customize a software application for determining sanitation suitability, embedding existing resources and reference data. Additional software applications could be used by decision-makers to obtain substantiated sanitation cost estimates upon adding **CONTEXT-SPECIFIC ADAPTATIONS** (e.g., types and quantities of materials) and local unit prices. Users could **REFERENCE A DATABASE** of similar cities to fill in cost data gaps. The application would 1) combine secondary datasets, satellite imagery, and user-input location-specific data to define neighborhood zones and their suitability for differing sanitation interventions, and 2) calculate city-wide sanitation costs using the combination of neighborhood-level interventions. Outputs would include **DOWNLOADABLE DATA AND MAPS.** The World Bank developed a beta version CWIS Costing and Planning Tool that helps users compare the capital and operation costs of different technical options (both onsite and offsite) at the component, system, or city level.[143]

5. HEALTH: ANTICIPATING WATERBORNE DISEASE OUTBREAKS

KEY INFORMANT

- Mayank Midha, Director, **GARV Toilets**, a social enterprise in India

BACKGROUND

Waterborne disease may be caused by ingestion or contact with viruses (e.g., rotavirus, Hepatitis E), bacteria (e.g., cholera, typhoid), protozoa (e.g., cryptosporidium, giardia), or helminths (e.g., schistosomiasis, Guinea worm). Collectively, waterborne pathogens infect millions of people every year, predominantly causing diarrheal disease in low-income countries among children under five.[144] Waterborne disease outbreaks can largely be prevented by widespread use of well-maintained WASH infrastructure, programs, and practices.[144] Through the 2005 International Health Regulations, WHO member states agreed to proactively protect global health security by detecting, assessing, and reporting public health events, including some waterborne disease outbreaks.[145] With the proliferation of electronic monitoring and reporting systems and internet-connected point-of-care devices, disease data should become increasingly available, even in low-resource contexts.[146]

WASH service provision and medical countermeasures (e.g., vaccines, rehydration therapy, antibiotics) continue to play important roles in reducing the negative impacts of diarrheal and other waterborne diseases. Innovative computational public health strategies, namely predictive disease surveillance and outbreak analytics, are becoming increasingly relevant and can complement these interventions.[147–152] In recent years, researchers have advanced understanding of the relationships between environmental conditions and waterborne disease occurrence and refined capabilities to project disease risks spatially and temporally.[153–155] COVID-19 models relevant to respiratory disease, hospitalization, and mortality forecasting have also proliferated[156] and could potentially be modified for other diseases.

Those responsible for preventing and mitigating waterborne disease outbreaks include national and local governments (e.g., ministries of health; regional or district public health offices), multilateral public health agencies (e.g., WHO) and implementing organizations (e.g., CSOs). These health professionals likely have varying degrees of technical expertise and availability, and would need to work closely with data scientists to develop models, conduct analyses, and interpret results. Comprehensive and accurate waterborne disease data would support public health decisions among multiple stakeholders, for example:

- Multilateral and bilateral development agencies seeking to disseminate best practices for data collection and early warning systems to help national and local governments prioritize public health responses.
- National governments developing intervention policies that address pressing public health and economic concerns.

- Local governments planning disease prevention interventions that engage and support vulnerable communities and populations.
- CSOs allocating and deploying resources (e.g., funds, equipment, personnel) quickly to the areas of greatest need during an outbreak.

NEEDS

Though substantial progress has been made in expanding coverage of WASH services, billions of people globally remain without access.[157] In low-income countries, even where safe water and sanitation services exist, they are not always used or delivered consistently.[158–160] As urbanization continues to crowd cities in the Global South, already strained WASH services may struggle to keep pace.[161,162] Natural disasters, conflicts, and climate change further exacerbate the risk of outbreaks.

Traditionally, cases identified in healthcare settings are confirmed with laboratory testing. Decision-makers decide when and where to intervene only after an outbreak has been confirmed. In low-resource settings, risks are elevated owing to fewer clinical laboratories and limited pathogen testing capacity. The public health community has long recognized the potential of turning the disease surveillance paradigm from that of "identification and response" to "prediction and prevention."[163,164] One interviewee noted:

> "With population increase in urban areas and increased dependency on shared sanitation services/facilities, we are moving in the direction of providing data to forecast disease outbreaks and hygiene parameters for strategic sanitation interventions."[165]

Limited data on pathogen occurrence, fate and transport, transmission, infectivity, and removal by water treatment processes, however, introduce uncertainty into predictive modeling.[166,167] Language and cultural barriers can also affect data literacy and harmonization.[168]

SOLUTIONS

Public health authorities could use prospective (i.e., modeled) trajectories of waterborne disease risks to effectively and efficiently allocate resources for outbreak prevention and control. The following **strategies** and solution examples address barriers by meeting aforementioned **USER REQUIREMENTS**:

Draw from parallel platforms for disease monitoring, visualization, and forecasting. Like waterborne diseases, vector-borne diseases such as malaria have climate and environmental drivers. Existing models include VectorMap[169] (a platform for visualizing vector distribution), CHIKRisk[170] (GLOBAL climate-based risk maps for chikungunya fever), Epidemic Prognosis Incorporating Disease and Environmental Monitoring for Integrated Assessment[171] (EPIDEMIA; a malaria **FORECASTING** system in Ethiopia), and Artificial Intelligence in Medical Epidemiology[172] (AIME; a Malaysian center for predicting trends in diseases such as dengue fever). The Global Water Pathogen Project, **HEALTHY FUTURES** Atlas, and Global Atlas of Helminth Infections also offer pathogen MAPS that could serve as useful examples for waterborne disease risk mapping.[173–175]

Pilot a new app for waterborne disease prediction. After gathering available data for a pilot location (e.g., one country), a research team could run spatiotemporal models that clarify predictable **RISK FACTORS** for diseases spread through well-understood transmission pathways. This information could be converted to a running application that shares and **INTERPRETS** disease risk maps (including model uncertainty) for decision-makers on a monthly or quarterly basis. Risk reports could be housed on an **OPEN** online platform, or remain privately accessible to relevant parties. Involving decision-makers in model development, upgrades, and **TRAINING** opportunities would be beneficial. Curated datasets of the model inputs could also be stored for reuse by other researchers. Potential open-access data sources include national disease surveillance systems, the Global Enteric Multicenter Study (GEMS),[176] Global Infectious Diseases and Epidemiology Online Network (GIDEON),[177] WASH coverage maps or surveys, and satellite and GIS data on land use, roads, healthcare facilities, climate, population density, and/or poverty.[178] Implement early warning systems. The Famine **EARLY WARNING** System Network (FEWS NET) funded by USAID applies a "prediction and prevention" paradigm to food insecurity.[179] This model could ultimately be extended to preventing waterborne disease.

6. Programming: Interpolating household data

KEY INFORMANTS

- Alberto Wilde, Country Director, **Global Communities**, Ghana
- Sara Marks, Professor, **Swiss Federal Institute of Aquatic Science and Technology (Eawag)**
- Zaituni Kanenje, Pro-Poor Manager, **Nakuru Water and Sanitation Services Company Limited (NAWASSCO)**, Kenya

BACKGROUND

Improved understanding of how demographic and WASH indicators vary within sub-national areas is increasingly recognized as important to targeting interventions and meeting the SDGs.[180] Some variables, such as poverty, population, and education, may inform multiple development goals.

Detailed spatial information on demographic, socioeconomic, environmental, and health indicators in low-resource settings is usually aggregated at regional or national levels. A few large-scale representative household surveys, such as DHS, MICS, and the World Bank's Living Standards Measurement Study (LSMS), collect data on regional and national trends. In some countries, district-level data are also available, coming from sources such as national censuses and other data initiatives like Open Development Cambodia or the District League Table in Ghana. High-resolution data maps are also becoming more available with expanded data extraction from satellite imagery. Using advances in spatial modeling methods, researchers have generated high-resolution maps by interpolating (predicting indicator values for unsampled locations) household survey data.[42,181–184] One example of this approach combined environmental data and mobile phone usage records to predict and map household poverty levels in Senegal.[42]

National governments (e.g., ministries of health), implementing organizations (e.g., CSOs), and researchers or evaluators most often require high-resolution (e.g., household-level) WASH data. These decision-makers might use the data to:

- Prioritize allocation of funding or program implementation timelines.
- Develop sound proposals within a short timeframe, limiting new data collection.[111]
- Tailor intervention selection or design to the local context.
- Evaluate how outcomes compare among projects.
- Track progress toward development goals.

NEEDS

Recent studies (as well as anecdotal experience in the field) emphasize the importance of highly localized contextual factors on program performance and sustained behavior change.[186,187] One CSO professional[185] noted;

> “for any program implementation, initial community assessment is critical.”

The need for public access to high-resolution, geo-referenced data on contextual factors is increasing, although field data collection remains resource intensive. One interviewee[188] noted;

> “Interpolated maps that can show the level of indicators in a region will be useful and timesaving when targeting interventions because data collection is time intensive and expensive. [This type of] tool will also provide a good understanding of the reality of a situation…”

Another noted the challenge of finding current household data;[116]

> “In most cases this kind of data is not updated frequently and government institutions depend on census data that is outdated based on when it was collected…. There should be a system for continuous consultation and update of data for evidence-based decision-making.”

Many organizations and researchers collect household data for their own program purposes; however, these data are not often available to the public.[189] Some programs treat data as confidential to avoid revealing implementation gaps or deficiencies that could affect funding opportunities. Further, consistent indicators are not necessarily used across projects, making it difficult to share information.[189] For example, WASHPaLS, a USAID-funded project, found rural sanitation implementation data to be highly inconsistent in coverage and quality.[190] In reviewing datasets from six programs covering a combined 40,482 communities in Cambodia, Ghana, Liberia, and Zambia, more than half of the data was rejected as potentially inaccurate during cleaning. Additionally, different programs used different reporting metrics to measure progress towards sanitation objectives. While some coordination mechanisms aim to combine results and lessons learned, these are notoriously difficult to coalesce at a granular level and not necessarily influential.[19,191]

Spatial interpolation methods are commonly used for elevation, temperature, precipitation, and soil mapping, but they have not commonly been applied to WASH data. This is due in part to model complexity and sparse data availability in low-resource and rural areas, where maps of even common variables are hard to come by. Other technical barriers include prediction uncertainty, differences in spatial resolution between urban and rural areas, and standardizing spatial resolutions and geographic extents among different data sources. Variables that are aggregated too broadly, and thus have low within-indicator variability, can cause a model to become unstable and affect the reliability of results. Ensuring reliable, comparable data requires clear rules and standards for data collection as well as capacity building for enumerators or users who input data.

SOLUTIONS

The goal of offering granular insights about household and community characteristics requires better leveraging of data collection efforts and increased support for user-friendly modeling applications. The following **strategies** and solution examples address barriers by meeting aforementioned **USER REQUIREMENTS**:

Enhance accessibility of DHS Modeled Map Surfaces for WASH actors. DHS Modeled Map Surfaces provide a standard set of spatially modeled maps from recent population-based surveys.[181] Each is produced

(cont.)

using standardized geostatistical methods, publicly available DHS data, and a standardized set of **PARAMETERS** across countries. Map packages contain a mean estimate surface, an **UNCERTAINTY** surface, and corresponding information on the model creation process and validation. In WASH applications, researchers observed disparities in latrine coverage among different wealth categories in Kenya using spatial interpolation.[192] Others have created **HIGH-RESOLUTION** poverty maps for countries including Kenya, Malawi, Nigeria, and Tanzania using spatial interpolation.[193]

» **Aggregate and harmonize data from multiple household surveys.** In addition to existing national surveys, smaller actors could leverage survey data by combining it with other data from the same program regions. Projects such as the Central American Rural Water and Sanitation Information System (SIASAR) and the Water Point Data Exchange (WPDx), for example, **COMPILE** data from multiple sources. SIASAR serves as a central, **OPEN** platform where governments and CSOs can upload and share WASH data, encompassing data collection, **VALIDATION**, analysis, and use through a suite of web-based apps.[194] WPDx, developed in partnership with businesses, CSOs, the World Bank, UNICEF, and World Vision, allows users across the globe to upload data on water points.[195] Using a set of rules to help **STANDARDIZE** data, WPDx aggregates information and allows users to explore data from about 250,000 water points in 25 countries.

» Develop a new web-based app that allows users to enter their own data and produce interpolated maps of high-resolution WASH indicators. Aquaya proposes an app that would emphasize user-friendliness and a **STEP-BY-STEP** approach, beginning with data formatting, **CHOOSING** an interpolation method, and selecting suitable covariates. Model fit and validation parameters would be run on the back end. The app would also produce a map of model uncertainty. **BASIC EXPLORATORY ANALYSES** would facilitate cluster-level observations and produce histograms and variograms. Guidance documentation geared toward program implementers can help to **ENHANCE CAPACITY** by providing use examples as well as survey design and data management recommendations. Users would be encouraged to SHARE their results to the Humanitarian Data Exchange, an open platform for disseminating data across organizations.

7. Programming: Understanding local contexts through community classification

KEY INFORMANTS

- Aaron Salzberg, Director, **The Water Institute at University of North Carolina**, Chapel Hill (UNC)
- Alberto Wilde (Country Director) and Eduardo Perez (Technical Director, Global WASH and Health), **Global Communities**
- Aliocha Salagnac (Information Management Systems Specialist), Michael Gnilo (Sanitation and Hygiene Specialist), and Rob Bain (Statistics and Monitoring Specialist), **UNICEF**
- Bal Mukunda Kunwar, Business Development Officer, **Helvetas Swiss Intercooperation Nepal**
- Carolien Van der Voorden (Head, Technical Support Unit) and Matteus Van der Velden (Manager, Asia Regional Unit), **Water Supply and Sanitation Collaborative Council (WSSCC)**
- Chris Cormency, Chief of WASH, **UNICEF Mozambique**
- Christopher Kanyesigye, Research and Development Manager, **National Water and Sewerage Corporation (NWSC) Uganda**
- Joseph Githinji, General Manager (Operations) and Senior Manager (Customer Support), **Sanergy**
- Kristoffer Welsien (Senior Water Supply & Sanitation Specialist) and Susanna Smets (Senior Regional Water Supply and Sanitation Specialist), **World Bank, Water Global Practice**
- Mbaye Mbeguere (Senior WASH Manager and Urban Focal Point), **WaterAid Senegal**

BACKGROUND

Improving the effectiveness of rural community interventions is critical to meeting SDG 6.2 targets (universal, adequate, and equitable sanitation access) and protecting public health. Community-Led Total Sanitation (CLTS) is the most widely used rural sanitation intervention in low-resource settings. WASH professionals increasingly acknowledge that different community contexts call for different approaches to encourage construction and sustained use of improved latrines. In the CLTS Handbook, Kar and Chambers recommend assessing communities for their "challenge level" (based on community characteristics) prior to initiating CLTS interventions.[196] For example, the Handbook states that more favorable outcomes are expected if communities are small and remote, with wet conditions, a high incidence of diarrhea, and no previous sanitation subsidy programs. The "Rethinking Rural Sanitation" Guidance (RRSG), developed by UNICEF, Plan International, and WaterAid, similarly calls for situational analyses at both national and district/province levels to determine community typologies and guide design of rural sanitation programs.[187]

National governments (e.g., ministries of water and sanitation), local agencies (e.g., public health offices), implementing organizations (e.g., CSOs), and funders (e.g., foundations, multilaterals) are typically tasked with targeting rural sanitation interventions. These decision-makers likely have some familiarity with interpreting data, statistics, and maps, but could augment capacity and save time by using software application outputs and visualizations. With access to community classification data, the following types of decisions could be simplified:

- Performing a rapid community assessment to determine which rural sanitation intervention(s), should be implemented in a given community, depending on the costs and likelihood of success.
- Prioritizing which communities are likely to benefit most from sanitation interventions.
- Detecting patterns in past community development efforts and outcomes that could be used to adapt and improve future sanitation interventions.
- Combining community typologies with additional data variables (e.g., precipitation, flooding) to further describe communities (e.g., by summarizing vulnerability to climate change).[21,54,91,197–201]
- Coordinating multiple WASH development organizations working in the same region.
- Informing national planning, guidance, and policies with up-to-date progress on community-level sanitation.
- Monitor overall progress toward global sanitation goals.[87]

NEEDS

Although the population in sub-Saharan Africa with basic sanitation services doubled from 2000–2017, estimates suggest that achieving universal coverage by 2030 will require further acceleration in expanding basic services.[69] Challenges are particularly stark in rural areas, where seven out of ten people lacked basic sanitation services as of 2017. Backsliding further threatens development investments and gains, as sustaining sanitation facilities and open-defecation-free (ODF) status over time has remained elusive. One interviewee[185] reported,

> “It's a circle that we see over and over happening because community assessment is not done properly from the beginning.”

The RRSG stops short of providing detailed methodologies for obtaining, integrating, or analyzing the necessary data to determine community typologies and the corresponding ideal programmatic mix.[187] Descriptive district-level data may be scarce due to the lack of incentives for data sharing. More recently, quantitative research by Aquaya has demonstrated the potential of using existing datasets on socio-economic and environmental characteristics to identify local contextual factors that influence variable CLTS program outcomes.[190] Nevertheless, data analysis apps relevant to rural sanitation program planning and design are in the early stages. For example, WaterAid developed a community-level app to assign community typologies; however, it requires surveying a community representative and is tough to apply at scale. Apps that do not depend on field-based data collection would

> “really help to solve challenges with regards to assessing large geographic areas.”[202]

Historical information about the successes or failures of previous sanitation interventions is also challenging to track, which may lead to repeating ineffective approaches.[185,203,204] One interviewee[197] noted,

> “Due to different stakeholder approaches on CLTS implementation and lack of data on what has been done, it becomes difficult to ‘trigger’ communities' behavior.”

Interviewees also described the challenges of planning and executing ongoing sanitation programming without duplicating efforts among multiple organizations:

> “[A] pain point is understanding what is happening on the ground during program implementation, especially when managing different partners. There is sometimes not enough data to really understand who is doing what and where.”[20]

SOLUTIONS

Providing decision-makers with community readiness classifications for sanitation and other types of WASH interventions will require evidence-informed data collation and processing approaches. The following **strategies** and solutions could help to meet the aforementioned **USER REQUIREMENTS:**

Leverage available data sources related to sanitation development goals. Sustainable Development Solutions Network has developed an **INTERACTIVE** website providing visual representation of all UN countries' SDG performance.[205] In addition to a **GLOBAL** map, the dashboard includes interactive plots and SDG indicator tiles. JMP has released an interactive web application to visualize national water and sanitation survey data.[206] Interactive features include the ability for users to build their own plots, download data, and export plot images. AtlasAI, a technology startup, develops data layers using remote sensing and AI, for example to generate detailed insights on poverty, crop yield, and economic trends across Sub-Saharan Africa.[207] **HIGH-RESOLUTION**, continental-scale map layers are offered as data downloads or as an interactive web map. The Socio-Economic Atlas of Kenya provides a **VISUAL** illustration of the geographic and socio-economic data pulled from the national census.[208]

Develop a single online platform with granular, geo-referenced data on community classifications. Following a desk review of existing sanitation tailoring guidance, Aquaya is working with stakeholders to a) pilot initial data variables (i.e., distance to roads, travel times to cities, and distance to towns) for calculating **COMMUNITY TYPOLOGIES** in select locations, b) create an interactive online platform, and c) test user functionalities. As part of the USAID-sponsored Water, Sanitation, Hygiene Partnerships and Learning for Sustainability (WASHPaLS) project, approximately 13 publicly available datasets, including the newly developed community typologies, have been added to a single, **USER-FRIENDLY** sanitation planning app called the Sanitation Planning Tool (SanPlan; Figure 2).[209] Covering more than 10 countries, SanPlan will offer five **ANALYSIS FEATURES** to help users explore highly localized (at least 5 km) spatial data. It can be expanded to include **MORE GEOGRAPHIC AREAS AND VARIABLES** and tested with structured feedback from potential users throughout the development process. The final application should allow users to modify rules, thereby employing their expert knowledge of CLTS performance and local context to **CUSTOMIZE** community classifications. Once launched, **PUBLIC TRAINING** events could be offered to aid dissemination.

Produce new data for potentially relevant sanitation success factors. Other important sanitation considerations identified in the CLTS Handbook, the RRSG, and a USAID desk review may require suitable **NEW DATASETS**.[186,187,196] These include income level/poverty, waterborne disease incidence, education level, favorable hydrogeology (water table level, soil conditions), involvement of local and traditional leadership, prior WASH and subsidy programming, gender equity in decision-making, road type (paved, all-weather, dirt, etc.), prevalence of agricultural livelihood, and prevalence of rented accommodation.

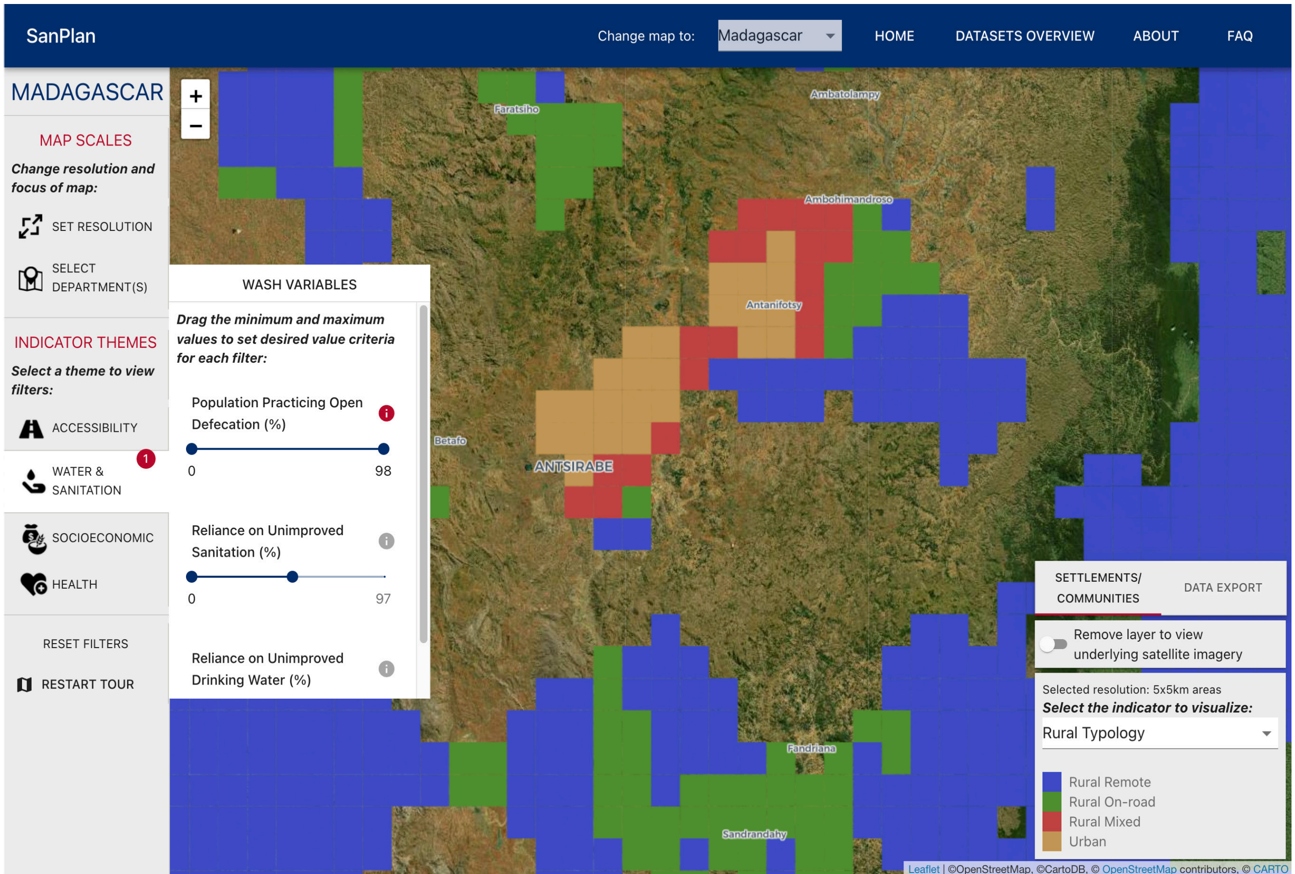


FIGURE 2: USAID-funded Sanitation Planning Tool (SanPlan, prototype under development) showing rural typology overlain with sanitation coverage in one area of Madagascar.

8. Programming: Targeting the poor and vulnerable

KEY INFORMANTS

- Chris Nicoletti (Senior Director of Impact and Analytics) and Abdul-Mumin Damba Tahidu (Country Manager, Ghana), **iDE**, an international CSO
- Elizabeth Tilley, Professor, **Malawi Polytechnic University**
- Issifu Adama, WASH Officer, **UNICEF Ghana**
- Mbaye Mbeguere, Senior WASH Manager and Urban Focal Point, **WaterAid Senegal**
- Philip Oyamo, Senior Project Manager, **WSUP Kenya**

BACKGROUND

Though access to safe WASH services in low-income countries has increased substantially over the past twenty years, governments must strive to ensure that poor and vulnerable households benefit equally.[210] The SDG imperative to "leave no one behind" requires dedicated financing strategies to improve water and sanitation access among the poorest two quintiles (40%) of the population. Loans may be appropriate for households that can afford to improve their WASH infrastructure when payments are spread out over time, while subsidies should be directed at households that simply cannot afford safe water and sanitation services (particularly when the costs exceed 5% of income). Institutions offering WASH financial support require indicators, preferably at a local level, to differentiate who should qualify. Potential data sources include nationally representative surveys (e.g., DHS, LSMS, census data), customer databases of micro-finance institutions, and satellite/aerial imagery (e.g., OneAtlas, Digital Globe).

The professionals responsible for identifying households that most need financial support include: national and local governments (often supported by donor and UN agencies), public or private service providers (e.g., utilities), CSOs, and funders (e.g., foundations). Within local governments, field staff have more experience collecting household data, while office staff have more expertise reviewing data summaries, statistics, and maps.

Common decisions that require localized poverty data (at the household or community level) include:

- Deciding which households would be able to repay a loan for infrastructure improvements such as a toilet or piped water connection.
- Determining which households require subsidies to make their water or sanitation service fees affordable.
- Allocating other social benefits (e.g., reduced-price health insurance, school fees, electricity, or fertilizers).
- Distributing aid in emergencies.

NEEDS

Organizations such as UNICEF and WaterAid have called for data-driven approaches to cost-effectively identify the poor without intensive additional data collection.[202,211] A Kenyan CSO working in urban Kenya reported,

> “Some of the decisions we have to make [are] identifying the correct communities to work in and the correct low-income areas. You would expect that utilities have full information on the areas they supply, but it's not the case.”[91]

This data paucity stems in part from constantly changing income levels and rapid urban expansion into new residential areas.

Robust methods for service providers or CSOs to identify the poorest households or communities are lacking. Pricing subsidies are common in the WASH sector, but they are poorly targeted and thus largely ineffective; in fact, they primarily benefit high-income groups and often fail to reach the very poor.[212,213] For example, block tariffs (setting pricing by a range of water usage) have been ineffective in low-income countries because utility services may exclude the poorest areas, and poor households often use more water to support large household sizes and/or their neighbors' connections.[214,215] Systems that rely on poverty criteria defined at the national level[216] (i.e., combining rural and urban areas) do not always provide sufficient sensitivity and differentiation at the local level.[217] Some rural WASH programs have used participatory input from community members to identify the poorest households,[132,218] but this resource-intensive approach is difficult to scale up quickly.[219] In addition, relying on few, more powerful individuals to decide which households qualify could introduce bias and corruption.[115] Other data sources such as high-resolution satellite imagery may be unavailable or outdated.

Loans for WASH upgrades are likewise increasingly prevalent, but lending institutions face difficulty in correctly identifying households that can afford to repay them.[220–222] Credit records are expensive to obtain and often inappropriate for evaluating borrowers in low-income countries who have limited-to-no credit history. A CSO offering loans for pour-flush toilets in northern Ghana found that existing credit-scoring software did not accurately predict repayment behaviors among their customers.[223] Customers' existing financial data may be difficult to access as they are often proprietary or protected by privacy laws.

SOLUTIONS

Increasing broad access to high-resolution information on poverty levels and creditworthiness requires more accurate, cost-effective, and scalable methods. The following **strategies** and solution examples address barriers by meeting aforementioned **USER REQUIREMENTS**:

Predict poverty status or credit score using “smart” surveys of simple household characteristics. A number of research studies have applied AI methods such as machine learning to determine whether a household is poor or not poor using **SIMPLE** survey data.[39,224,225] World Bank researchers have demonstrated this approach in Malawi and Indonesia and are incorporating AI into **CONCISE**, dynamic poverty identification surveys.[226] The USAID Indonesia Urban Water, Sanitation and Hygiene (IUWASH) PLUS program in Indonesia showed that a short, **RAPID** survey provided sufficient information to map poverty clusters at the city level.[227] For loan

(cont.)

evaluation, machine learning could identify a small number of customer characteristics that **ACCURATELY** predict loan repayment. An organization providing WASH loans would need to pilot the novel smart survey approach over 1–2 years to validate, test, and refine the survey application.

Use remote sensing data to map poverty. Studies have shown that satellite imagery (particularly night images) and mobile phone usage data can be used to map poverty at the sub-national level.[40,47,228] Facebook Data for Good, which uses social media data, and AtlasAI offer high-resolution (about 2 km) relative wealth maps.[207,229] Researchers have also applied satellite imagery to identify poverty levels at very HIGH **RESOLUTION** (household level or 100 m), although the performance of these models could be improved.[230] PulseSatellite, a collaborative satellite image analysis tool using **HUMAN-AI FEEDBACK LOOPS**, has been evaluated in WASH-related humanitarian contexts for mapping refugee settlements and flooded areas.[231]

Develop a mobile app that merges these technologies to assign poverty status. If census data is not available, household surveyors could administer a short, "smart," mobile survey consisting of a maximum 10 questions selected to most accurately identify poverty (and **VALIDATED** against other methods). An application could then use an embedded **PREDICTIVE** model to rapidly generate the household's status, so the surveyors might provide a "poverty card" to households that qualify in real time. In urban settings, it may be more realistic to limit the survey to a representative sample of households within similar zones. Data could be **AUTOMATICALLY TRANSMITTED** to populate a poverty map. Upon completion, satellite or aerial imagery could be merged with the map, and an algorithm could be **TRAINED TO RECOGNIZE** impoverished areas using artificial intelligence. This would provide a simplified approach to map updates, although analysts would need to ground-truth the algorithm in a few locations and renew the imagery at least once a year.

9. Programming:
Evaluating impacts

KEY INFORMANTS

- Antoinette Kome, Global Sector Coordinator WASH, **SNV** (Netherlands Development Organisation)
- Bal Mukund Kunwar, Business Development Officer, **Helvetas Swiss Intercooperation Nepal**
- Carolien Van der Voorden, Head of Technical Support Unit, **WSSCC**
- Isaac Kega, Monitoring and Evaluation Specialist, and Stella Warue, Programme Officer, **Water Sector Trust Fund**
- Zaituni Kanenje, Pro-Poor Manager, **NAWASSCO,** Kenya

BACKGROUND

Collecting rigorous evidence to measure WASH program impacts remains complex for many reasons, including the multifaceted nature of interventions and the influences of external factors (confounders) on program outcomes. To measure impact, researchers typically rely on a control group, which provides a "counterfactual" to the intervention, to understand what might have happened if no intervention was applied. Randomized control trials are considered optimal for comparing control vs. treatment groups, but these studies are costly, time-consuming, and sometimes unethical or unrealistic in development settings.[232] Observational study designs that measure outcome trends over time (and may include comparisons between groups that did and did not receive the intervention) offer alternative evaluation approaches, albeit more susceptible to confounders.

Evaluation studies typically produce results near the end of WASH projects or following completion. However, real-time program data combined with emerging analytical approaches could provide opportunities to adjust or calibrate interventions prior to program completion. For example, machine-learning techniques could be applied to available datasets to identify matched control groups and track longitudinal data (change over time).

Those responsible for evaluating WASH intervention outcomes and maximizing impacts include national governments, implementing organizations (e.g., CSOs), and donors. Engendering partnerships between WASH practitioners and data science experts could benefit decision-making,[233] for example to:

- Decide which WASH programs should be continued, scaled up, replicated, or discontinued.
- Assess whether a WASH investment or program is on track to reach its intended goals and, if not, learn what barriers might be present.

- Determine whether WASH programs met expected goals and justify future funding requests.
- Compare costs and benefits across WASH interventions and set impact goals.
- Describe contextual factors that influence WASH intervention outcomes.
- Modify WASH interventions based on preliminary data on program impacts.

NEEDS

Among the complex landscape of WASH and other development initiatives, attributing impacts to a specific project is challenging. Health impacts in particular have been difficult to demonstrate; of 10 randomized control trials studying sanitation interventions, only three found statistically significant reductions in diarrheal disease.[12] A systematic review of the CLTS methodology similarly found limited evidence of sustained sanitation behavior change, despite its implementation in more than 50 countries.[234,235] Donors desire clear evaluation data to guide investments in effective programs. One interviewee[87] noted,

> “We struggle to make a case for additional investments, because it is hard to know how investments in water and sanitation lead to economic development.”

Although large implementing organizations may have more capacity to design proper monitoring and evaluation systems for their programs, “success” remains subjective. One CSO interviewee[236] mentioned,

> “Most donors, even implementors, don’t have a realistic idea about what reasonable success could even be for a program.”

Smaller organizations struggle even more to evaluate their efforts. A manager[116] focused on pro-poor access at an urban water utility noted,

> “[We] have a challenge on the kind parameters to use and measure the impact and sustainability of [our] programs. [We] do not know the kind of data that needs to be collected and how to evaluate it.”

Rural WASH program implementors[197] also require improved evaluation systems:

> “[Our] greatest need is on assessing the performance of community-based institutions supporting water interventions and what kind of data are useful to measure the performance of rural WASH programs.”

SOLUTIONS

Attaining speedier insights into WASH intervention outcomes requires advanced methods that take greater advantage of existing data and prioritize new data collection. The following strategies and solution examples address barriers by meeting aforementioned **USER REQUIREMENTS:**

Gather existing, “passively” collected data. Ideally, datasets should include intervention and non-intervention locations or pre-implementation baseline conditions. Potential data sources include large-scale national surveys, user-created data (e.g., on social media), transactional data (e.g., purchase information), GPS/location data, satellite imagery, although such datasets may not offer equitable representation.[233] The UN Millennium Development Campaign in partnership with the WSSCC and UN Global Pulse used **EXISTING** social media data to evaluate the effectiveness of a sanitation campaign.[237] It relied on natural language **PROCESSING TECHNIQUES** to

(cont.)

evaluate perceptions and sentiments about sanitation. Correlating spikes in social media posts with the timing of relevant sanitation campaigns (e.g., World Toilet Day) provided a clear baseline to monitor the effectiveness and reach of the communications campaign in **REAL TIME**.

Process datasets to obtain new insights. Local officials in Florida combined 80 different administrative and program datasets covering 80,000 child welfare cases over a period of five years to determine which actions and approaches result in the best outcomes for children.[233] They developed an **ALGORITHM** that matched cases in terms of background and context, while varying the types of "treatment" received by each child to understand which led to the desired impact. Specific to WASH, the Conrad N. Hilton Foundation's Safe Water program, evaluated by Stanford University, aims to increase access to safe water services in 12 focus districts.[238,239] Data on a number of hypothesized **INFLUENTIAL FACTORS** affecting access to safe water (e.g., population density, poverty, terrain) were gathered for both intervention and comparison districts to confirm their similarities. This reduced the need for extensive primary data collection outside of the planned program activities.

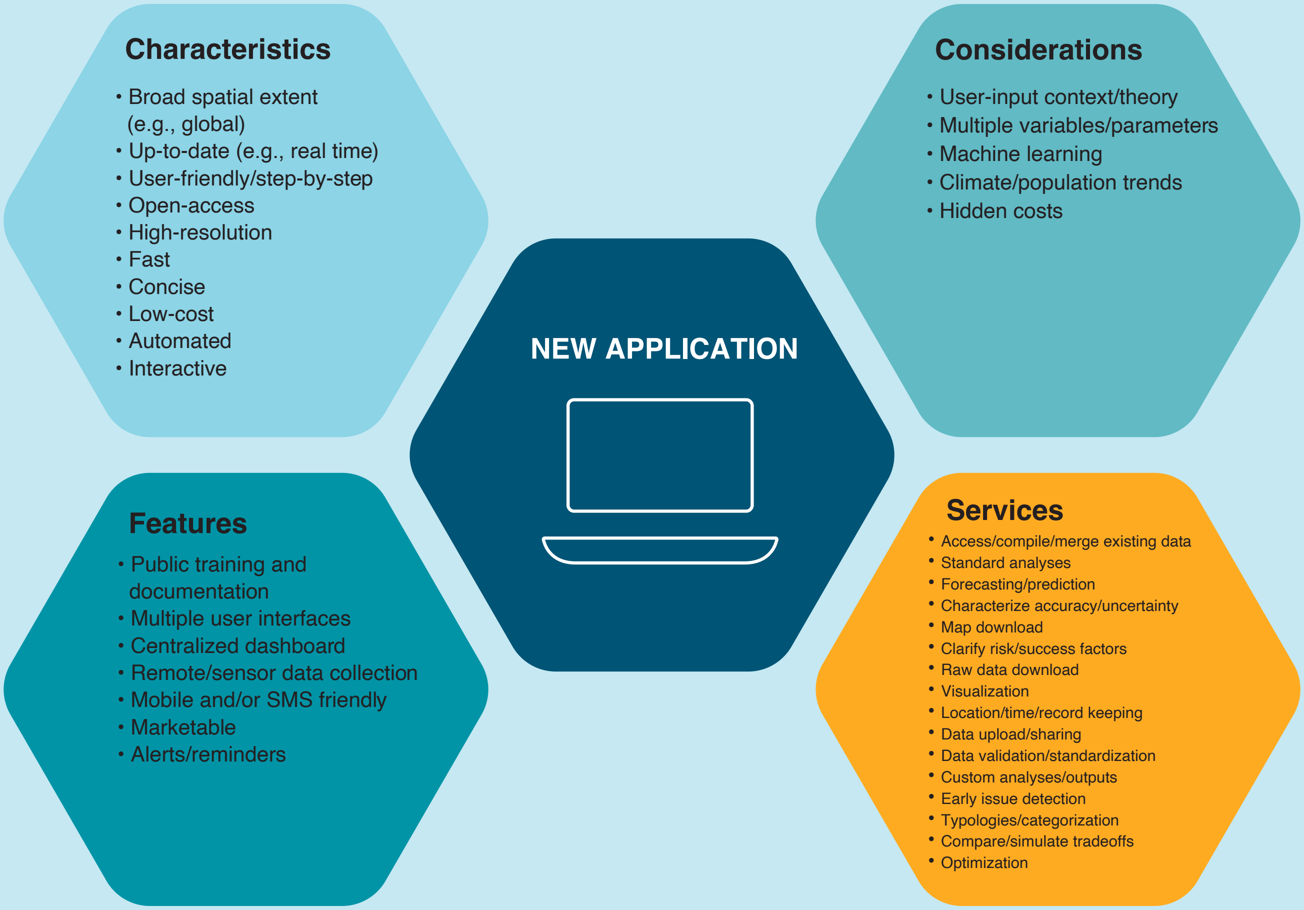
Test cases that further the vision of real-time WASH program evaluation. To avoid conflating correlation and causation, data science approaches for WASH program evaluation necessitate an iterative approach in collaboration with subject matter experts. This process should co-produce the underlying **THEORY OF CHANGE** and explanatory data model, identifying success variables that can be converted to **GOALS AND TARGETS** for future programs. Ongoing monitoring can be automated by integrating predictive analytics (for large prospective studies) or performance indicators into an online dashboard that would **ALERT** implementers or funders when projects are not on track to meet specific targets.

Conclusions:

Where are the opportunities for applying data science to improve WASH decisions?

The use case examples presented above offer insight into WASH needs and demands (Figure 3; Figure 4), each with the potential to fill an important gap faced by decision-makers. Yet, they vary in the status of development, demonstrated feasibility of use, and potential reach. Other WASH data science applications (beyond those featured here) may already be pertinent for various stakeholders or become more pressing in the future. **In the absence of unlimited funds and resources, determining which solutions to invest in requires assessments of the trade-offs between effort and potential impact.**

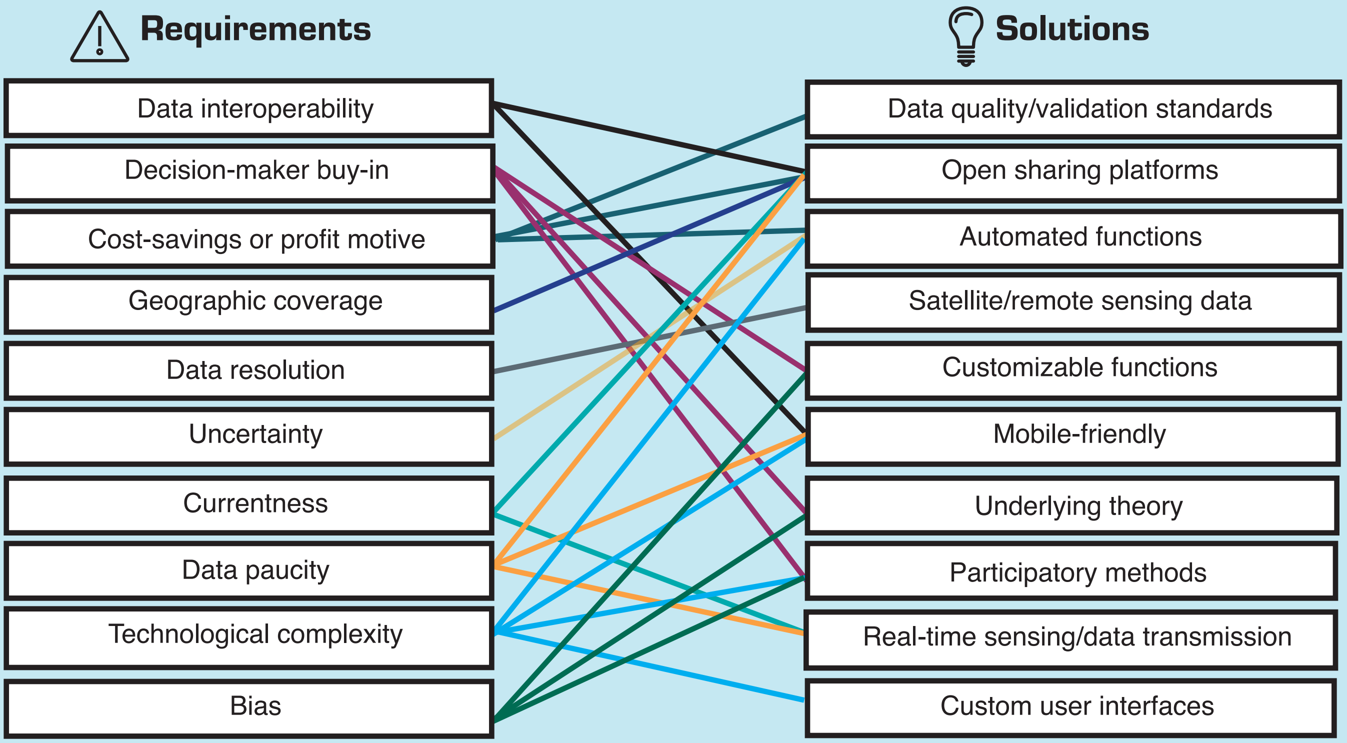
FIGURE 3: Technical requirements for new applications (or other data tools) gleaned from interviewees across the nine use cases.



WHICH USE CASES ARE MOST AND LEAST COMPLEX TO REALIZE?

Greater detail on next steps (proposed outputs, data sources, barriers, and workflows) for each featured use case appears in the companion annex, Annex 1: Analysis of Results by Use Cases. Possibly the most complex use case is “understanding sanitation costs.” The solution itself has many components that require disparate data sources, as well as a substantive need to garner understanding and input from local users; thus, it would require thoughtful research, guidance, and training. The use case “understanding local contexts through community classification” is probably the least complex, as it does not require comparatively sophisticated machine learning techniques to develop and is less dependent on user inputs. It is important to note, however, that the low complexity is not an indication of simplicity, rather comparative complexity to the other use cases.

FIGURE 4: Diagram connecting key requirements for WASH data science with potential solutions



WHICH USE CASES WOULD HAVE THE GREATEST IMPACT?

All use cases are expected to have a high impact. The use cases subjectively judged by the research team to have the highest potential impact are:

- “Targeting the poor and vulnerable,”
- “Anticipating waterborne disease outbreaks,”
- “Fecal sludge emptying needs,” and
- “Understanding sanitation costs.”

These use cases generally share examples of tools, applications, or methodologies from which to draw inspiration. In contrast, the use case for “reducing non-revenue water” is expected to have relatively less impact due to a number of similar emerging innovations and strong interest from other developers in the same space.

WHAT IS AQUAYA'S ROLE IN IMPLEMENTING THE USE CASES?

Aquaya has experience relevant to most of the featured use cases. Notably, the USAID-funded WASHPaLS project enlisted Aquaya to develop a web-based community classification tool to support rural sanitation development, as mentioned in the "understanding local contexts through community classification" use case. Aquaya also recently completed projects related to sanitation costing and latrine pit emptying in Kenya, Ghana, Tanzania, and Nigeria, groundwater quality modeling in Uganda and Bangladesh, and identifying poor and vulnerable populations for targeting sanitation subsidies in Ghana. Through numerous projects on these topics, Aquaya has developed an expansive network of partners in implementation, including local and national governments, water utilities, development organizations, and other stakeholders. Aquaya and other WASH research institutions are thus well-positioned to move forward with development of WASH data science projects.

WHICH ETHICAL CONSIDERATIONS INFLUENCE WASH DATA SCIENCE?

When decision-makers rely on data to make critical policy or investment decisions affecting individuals and communities, the data must be of good quality and representative of the population. This is important not only to meet a project or program's intended goals, but to ensure decisions are free from bias and discrimination and will not cause harm or unintended consequences. **Principles for ethical and responsible data collection and use are standard for most researchers, with the ultimate goal of protecting human research subjects.** Standard procedures include submission of surveys and data analysis protocols for review and approval by one or more independent Institutional Review Boards (IRBs), ethics committees, or research governance boards; obtaining informed consent or assent from participants; sharing and explaining research goals, methods, and findings to promote transparency; and anonymizing results as needed. Researchers should strive to represent the interests of both professionals and consumers from all walks of society, including low-income, vulnerable, and marginalized communities. In practice, this might manifest as ensuring research teams represent both the Global North and Global South, or targeting specific levels of participation by women or minorities.

While big data and data analytics offers efficiencies for research and opportunities to explore more complex datasets, they also open new risks of bias and non-representative data. Relying on existing datasets rather than primary data collection means that researchers and decision-makers may have less interaction with communities, do not hear about challenges firsthand, or neglect to ground-truth the theory underlying key analytical decisions.[233] Accelerating or automating data collection steps risks misinterpretation of results and promoting decision-making that is not inclusive. For instance, big datasets collected with mobile phones (or that rely on internet access) risk being non-representative of certain groups who may not have equal access to this technology, such as women or the poor.[233] As the volume of available data continues to expand and technologies advance, WASH professionals must not only continue to maintain standard practices for ethics in research, but make explicit efforts to mitigate potential risks of introducing advanced technologies across all steps of data management, including collection, analysis, storage, and destruction.[240]

Recommended Reading

Anderson, C. (2008, June 23). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Wired. https://www.wired.com/2008/06/pb-theory/

Bartram, J., & Setty, K. (2021). Environmental Health Science and Engineering for Policy, Programming, and Practice. Journal of Environmental Engineering, 147(10), 03121002. https://doi.org/10.1061/(asce)ee.1943-7870.0001880

Cantor, A., Kiparsky, M., Kennedy, R., Hubbard, S., Bales, R., Pecharroman, L., Guivetchu, K., McCready, C., & Darling, G. (2018). Data for Water Decision Making: Informing the Implementation of California's Open and Transparent Water Data Act through Research and Engagement. https://doi.org/10.15779/J28H01

McAfee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. https://hbr.org/2012/10/big-data-the-management-revolution

Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision-Making. Big Data, 1(1). https://doi.org/10.1089/big.2013.1508

Rice, M. (2019). 17 Big Data Examples and Applications. Builtin. https://builtin.com/big-data/big-data-examples-applications

WaterAid. (2019). From Data to Decisions: How to Promote Evidence-Based Decision Making Through External Investments in Country-Led Monitoring Processes. https://washmatters.wateraid.org/sites/g/files/jkxoof256/files/from-data-to-decisions.pdf

York, P., & Bamberger, M. (March 2020). Measuring Results and Impact in the Age of Big Data: The Nexus of Evaluation, Analytics, and Digital Technology. Rockefeller Foundation: New York. https://www.rockefellerfoundation.org/report/measuring-results-and-impact-in-the-age-of-big-data-the-nexus-of-evaluation-analytics-and-digital-technology/

References

1. Wolvers, A., Tappe, O., Salverda, T., & Schwarz, T. (n.d.). Concepts of the Global South – Voices from around the world.

2. Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision-Making. Big Data, 1(1). https://doi.org/10.1089/big.2013.1508

3. Hays, C. L. (2004). What Wal-Mart Knows About Customers' Habits. New York Times.

4. Anderson, C. (2008, June 23). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Wired. https://www.wired.com/2008/06/pb-theory/

5. Rice, M. (2019). 17 Big Data Examples and Applications. Builtin. https://builtin.com/big-data/big-data-examples-applications

6. Cantor, A., Kiparsky, M., Kennedy, R., Hubbard, S., Bales, R., Pecharroman, L., Guivetchu, K., McCready, C., & Darling, G. (2018). Data for Water Decision Making: Informing the Implementation of California's Open and Transparent Water Data Act through Research and Engagement. https://doi.org/10.15779/J28H01

7. McAfee, A., & Brynjolfsson, E. (2012, October). Big Data: The Management Revolution. Harvard Business Review.

8. Nilsson, N. J. (2009). Artificial Intelligence: A New Synthesis. Morgan Kaufmann Publishers, Inc.

9. Langley, P. (1996). Elements of Machine Learning. Morgan Kaufmann Publishers, Inc.

10. Bradbury, K., Saboo, R., Johnson, T. L., Malof, J. M., Devarajan, A., Zhang, W., Collins, L. M., & Newell, R. G. (2016). Distributed solar photovoltaic array location and extent dataset for remote sensing object identification. Scientific Data, 3(160106). https://doi.org/10.1038/sdata.2016.106

11. Prüss-Ustün, A., Wolf, J., Bartram, J., Clasen, T., Cumming, O., Freeman, M., Gordon, B., Hunter, P., Medlicott, K., & Johnston, R. (2019). Burden of disease from inadequate water, sanitation and hygiene for selected adverse health outcomes: An updated analysis with a focus on low- and middle-income countries. Journal of International Hygiene and Environmental Health, 222(5), 765–777. https://doi.org/10.1016/j.ijheh.2019.05.004

12. Whittington, D., Radin, M., & Jeuland, M. (2020). Evidence-based policy analysis? The strange case of the randomized controlled trials of community-led total sanitation. Oxford Review of Economic Policy, 36(1), 191–221. https://doi.org/10.1093/oxrep/grz029

13. Pickering, A. J., Null, C., Winch, P. J., Mangwadu, G., Arnold, B. F., Prendergast, A. J., Njenga, S. M., Rahman, M., Ntozini, R., Benjamin-Chung, J., Stewart, C. P., Huda, T. M. N., Moulton, L. H., Colford, J. M., Luby, S. P., & Humphrey, J. H. (2019). The WASH Benefits and SHINE trials: interpretation of WASH intervention effects on linear growth and diarrhoea. The Lancet Global Health, 7(8), e1139–e1146. https://doi.org/10.1016/S2214-109X(19)30268-2

14. United Nations. (2020). Goal 6: Ensure access to water and sanitation for all. Sustainable Development Goals. https://www.un.org/sustainabledevelopment/water-and-sanitation/

15. Setty, K., Cronk, R., George, S., Anderson, D., O'Flaherty, G., & Bartram, J. (2019). Adapting translational research methods to water, sanitation, and hygiene. International Journal of Environmental Research and Public Health, 16(20). https://doi.org/10.3390/ijerph16204049

16. WaterAid. (2019). From data to decisions: how to promote evidence-based decision making through external investments in country-led monitoring processes.

17. Interview with Duncan McNicholl, Director, Uptime. (2020).

18. Interview with Fiona Zakaria, Research Fellow, University of Leeds. (2020).

19. Kumpel, E., MacLeod, C., Stuart, K., Cock-Esteb, A., Khush, R., & Peletz, R. (2020). From Data to Decisions: Understanding information flows within regulatory water quality monitoring programs. Npj Clean Water, 3, 38.

20. Interview with Carolien Van der Voorden (Head, Technical Support Unit), Water Supply and Sanitation Collaborative Council (WSSCC). (2020).

21. Interview with Chris Cormency, Chief of WASH, UNICEF Mozambique. (2020).

22. Bartram, J., Brocklehurst, C., Fisher, M. B., Luyendijk, R., Hossain, R., Wardlaw, T., & Gordon, B. (2014). Global monitoring of water supply and sanitation: History, methods and future challenges. International Journal of Environmental Research and Public Health, 11(8), 8137–8165. https://doi.org/10.3390/ijerph110808137

23. UNICEF. (2021). Multiple Indicator Cluster Surveys (MICS). https://mics.unicef.org/

24. USAID. (2021). The DHS Program - Quality information to plan, monitor and improve population, health, and nutrition programs. 25. WHO. (2021). UN-Water Global Analysis and Assessment of Sanitation and Drinking-Water (GLAAS). https://dhsprogram.com/

26. Rahman, Z., Crocker, J., Chang, K., Khush, R., & Bartram, J. (2011). A comparative assessment of institutional frameworks for managing drinking water quality. Journal of Water, Sanitation and Hygiene for Development, 1(4), 242–258. https://doi.org/10.2166/washdev.2011.002

27. Chouler, J., & Di Lorenzo, M. (2015). Water quality monitoring in developing countries; Can microbial fuel cells be the answer? In Biosensors (Vol. 5, Issue 3, pp. 450–470). MDPI AG. https://doi.org/10.3390/bios5030450

28. Andres, L., Boateng, K., Borja-Vega, C., & Thomas, E. (2018). A review of in-situ and remote sensing technologies to monitor water and sanitation interventions. Water (Switzerland), 10(6). https://doi.org/10.3390/w10060756

29. CartONG, UNHCR, & Tdh. (2017). Benchmarking of Mobile Data Collection Solutions: What aspects to consider when choosing a tool/platform.

30. mWater Foundation. (2020). mWater: Mobile Technology for Social Water Monitoring in Low-resource Settings [mobile application software].

31. CommCare by Dimagi | Data Collection App. (2021). https://www.dimagi.com/commcare/

32. ODK Docs. (2020). ODK Central. https://docs.getodk.org/central-intro/

33. Magpi. (2021). https://www.magpi.com/

34. Akvo Foundation. (2021). Capture reliable data. https://akvo.org/capture-reliable-data/

35. Dharma Platform. (2021). Data management solutions for impact-first organizations. https://dharmaplatform.com/

36. Poimapper. (2019). Advanced Mobile Forms & Mobile Data Collection. https://www.poimapper.com/en/

37. Mobenzi. (2021). Elegant technology enabling impact. https://www.mobenzi.com/

38. World Bank Group, W. B. (n.d.). Measuring income and poverty using Proxy Means Test.

39. McBride, L., & Nichols, A. (2018). Retooling poverty targeting using out-of-sample validation and machine learning. World Bank Economic Review, 32(3), 531–550. https://doi.org/10.1093/wber/lhw056

40. Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. Science, 350(6264), 1073–1076. https://doi.org/10.1126/science.aac4420

41. Steele, J. E., Sundsøy, P. R., Pezzulo, C., Alegana, V. A., Bird, T. J., Blumenstock, J., Bjelland, J., Engø-Monsen, K., de Montjoye, Y.-A., Iqbal, A. M., Hadiuzzaman, K. N., Lu, X., Wetter, E., Tatem, A. J., & Bengtsson, L. (2017). Mapping poverty using mobile phone and satellite data. Journal of The Royal Society Interface, 14(127), 20160690. https://doi.org/10.1098/rsif.2016.0690

42. Pokhriyal, N., & Jacques, D. C. (2017). Combining disparate data sources for improved poverty prediction and mapping. Proceedings of the National Academy of Sciences of the United States of America, 114(46), E9783–E9792. https://doi.org/10.1073/pnas.1700319114

43. Blumenstock, J. E. (2016). Fighting poverty with data. Science, 353(6301), 753–754. https://doi.org/10.1126/science.aah5217

44. Noor, A. M., Alegana, V. A., Gething, P. W., Tatem, A. J., & Snow, R. W. (2008). Using remotely sensed night-time light as a proxy for poverty in Africa. Population Health Metrics, 6(1), 5. https://doi.org/10.1186/1478-7954-6-5

45. Mellander, C., Lobo, J., Stolarick, K., & Matheson, Z. (2015). Night-Time Light Data: A Good Proxy Measure for Economic Activity? PLOS ONE, 10(10), e0139779. https://doi.org/10.1371/journal.pone.0139779

46. Bruederle, A., & Hodler, R. (2018). Nighttime lights as a proxy for human development at the local level. PLoS ONE, 13(9). https://doi.org/10.1371/journal.pone.0202231

47. Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. Science, 353(6301), 790–794. https://doi.org/10.1126/science.aaf7894

48. Abelson, B., Varshney, K. R., & Sun, J. (2014). Targeting Direct Cash Transfers to the Extremely Poor. https://doi.org/10.1145/2623330.2623335

49. SweetSense. (2021). Internet of Broken Things.

50. Sanitation, W. J. M. P. (JMP) for W. S. and. (2021). Water quality monitoring. https://washdata.org/monitoring/drinking-water/water-quality-monitoring

51. Aquaya Institute. (2021). Project W. https://aquaya.org/project-w/

52. Interview with Adam Harvey (Managing Director) and Frédéric Bergeron (General Manager), Whave. (2020).

53. Interview with Roland Liemberger, Independent Non-Revenue Water Consultant. (2020).

54. Interview with Joseph Githinji, (General Manager, Operations and Senior Manager, Customer Support), Sanergy Kenya. (2020).

55. Interview with Robel Lambisso, Technical Program Manager, World Vision Ghana. (2020).

56. Interview with Harrison Kwach, Independent WASH Consultant, Urban Waste Management. (2020).

57. Bartram, J., & Setty, K. (2021). Environmental Health Science and Engineering for Policy, Programming, and Practice. Journal of Environmental Engineering, 147(10), 03121002. https://doi.org/10.1061/(asce)ee.1943-7870.0001880

58. Paynter, N. (2021). The Other Side Of the Data Horizon (Draft).

59. Setty, K., Jiménez, A., Willetts, J., Leifels, M., & Bartram, J. (2020). Global water, sanitation and hygiene research priorities and learning challenges under Sustainable Development Goal 6. Development Policy Review, 38(1), 64–84. https://doi.org/10.1111/dpr.12475

60. Ransom, K., Nolan, B., Traum, J., Faunt, C., Bell, A., Gronberg, J., Wheelerg, D., Rosecrans, C., Jurgens, B., Schwarz, G., Belitz, K., Eberts, S., Kourakos, G., & Harter, T. (2017). A hybrid machine learning model to predict and visualize nitrate concentration throughout the Central Valley aquifer, California, USA. Science of The Total Environment, 601–602, 1160–1172.

61. Anning, D., Paul, A., McKinney, T., Huntington, J., Bexfield, L., & Thiros, S. (2012). Predicted Nitrate and Arsenic Concentrations in Basin-Fill Aquifers of the Southwestern United States. US Geological Survey Scientific Investigations Report 2012-5065

62. Weber, M. C. (2015). Modeling groundwater quality in an arid agricultural environment in the face of an uncertain climate: the case of Mewat District, India. https://doi.org/10.17077/etd.c8twawh7

63. Kenda, K., Čerin, M., Bogataj, M., Senožetnik, M., Klemen, K., Pergar, P., Laspidou, C., & Mladenić, D. (2018). Groundwater Modeling with Machine Learning Techniques: Ljubljana polje Aquifer. Proceedings, 2(11), 697. https://doi.org/10.3390/proceedings2110697

64. Podgorski, J., & Berg, M. (2020). Global threat of arsenic in groundwater. Science, 368, 845–850. https://doi.org/10.1126/science.aba1510

65. IPCC. (2014). Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (V. R. Barros, C. B. Field, D. J. Dokken, M. D. Mastrandrea, K. J. Mach, T. E. Bilir, M. Chatterjee, K. L. Ebi, Y. O. Estrada, R. C. Genova, B. Girma, E. S. Kissel, A. N. Levy, S. MacCracken, & and L. L. W. P.R. Mastrandrea (Eds.)). Cambridge University Press.

66. Damania, R., Desbureaux, S., Rodella, A.-S., Russ, J., & Zaveri, E. (2019). Quality Unknown: The Invisible Water Crisis. World Bank: Washington, D.C.

67. Sherbinin, A., Castro, M., Gemenne, F., Adamo, S., Cernea, S., Fearnside, P., Krieger, G., Lahmani, S., Oliver-Smith, A., Pankhurst, A., Scudder, T., Singer, B., Tan, Y., Wannier, G., Boncour, P., Ehrhart, C., Hugo, G., Pandey, B., & Shi, G. (2011). Preparing for Resettlement Associated with Climate Change. Science, 334(6055). https://doi.org/10.1126/science.1208821

68. MacDonald, A., Calow, R., MacDonald, D., Darling, W., & Dochartaigh, B. (2009). What impact will climate change have on rural groundwater supplies in Africa? Hydrological Sciences Journal, 54(4), 690–703. https://doi.org/10.1623/hysj.54.4.690

69. WHO/UNICEF. (2017). Progress on Household Drinking Water, Sanitation and Hygiene 2000-2017: Special Focus on Inequalities.

70. Interview with Ramon Brentfüherer (Project Manager, Groundwater Policy Advice) and Vincent Post (Hydrogeologist), BGR. (2020).

71. Interview with Aude-Sophie Rodella (Senior Economist), Esha Zaveri (Economist), Jason Russ (Economist), and François Bertone (Senior Water Resource Management Specialist, Groundwater), World Bank, Global Water Practice. (2020).

72. Interview with Neno Kukuric (Director) and Claudia Ruiz Vargas (Researcher), IGRAC. (2020).

73. Kut, K. M. K., Sarswat, A., Srivastava, A., Pittman, C. U., & Mohan, D. (2016). A Review of Fluoride in African Groundwater and Local Remediation Methods. Groundwater Sustainable Development, 190–212. https://doi.org/10.1016/j.gsd.2016.09.001

74. World Business Council for Sustainable Development. (2019). India Water Tool 3.0.

75. Global Water Intelligence. (2021). https://www.globalwaterintel.com/

76. Kuzma, S., Kerins, P., Saccoccia, E., Whiteside, C., Roos, H., & Iceland, C. (2020). Leveraging Water Data in a Machine Learning–Based Model for Forecasting Violent Conflict.

77. International Groundwater Resources Assessment Centre (IGRAC). (n.d.). Global Groundwater Information System (GGIS).

78. Kingdom, B., Soppe, G., & Sy, J. (2016). What is non-revenue water? How can we reduce it for better water service? World Bank Blogs. https://blogs.worldbank.org/water/what-non-revenue-water-how-can-we-reduce-it-better-water-service

79. Kingdom, B., Liemberger, R., & Marin, P. (2006). The Challenge of Reducing Non-Revenue Water (NRW) in Developing Countries and How the Private Sector Can Help: A Look at Performance-Based Service Contracting.

80. PPIAF. (2020). Developing Good Performance Based Contract Practices for Non-Revenue Water Management.

81. IWA, & World Bank. (2016). The World Bank and the International Water Association to Establish a Partnership to Reduce Water Losses. https://www.worldbank.org/en/news/press-release/2016/09/01/the-world-bank-and-the-international-water-association-to-establish-a-partnership-to-reduce-water-losses

82. PPIAF. (2019). Public-Private Infrastructure Advisory Facility Annual Report 2019.

83. WASREB. (2018). Non-Revenue Water Audit of WSPs: Final Report.

84. PPIAF (source). (2021, November 18). Ethiopian contract makes a case for NRW work in Africa. Global Water Intelligence (GWI) Magazine.

85. Utilis. (2020). Hydro-ScanTM. https://utiliscorp.com/product/hydro-scan/

86. Asterra: Technology by Utilis. (2020). Case Studies: Bangkok, Thailand.

87. Interview with Isaac Kega (Monitoring and Evaluation Specialist) and Stella Warue (Programme Officer), Water Sector Trust Fund, Kenya. (2020).

88. Interview with Habab Taifour (Senior Water Resources Specialist), World Bank Ethiopia. (2020).

89. WASREB. (2016). Tariff Guidelines.

90. WSUP. (2017). A Guide to Non-Revenue Water Reduction: How to Limit Losses, Strengthen Commercial Viability, and Improve Services.

91. Interview with Philip Oyamo, Senior Project Manager, Water & Sanitation for the Urban Poor (WSUP) Kenya. (2020).

92. XiO. (2020). Solutions | XiO Water Systems. https://xiowatersystems.com/cloud-scada-solutions/

93. Metasphere. (2020). Water pressure and flow monitoring: more intelligent, efficient, and resilient networks. Metasphere: Make Data Count. https://www.metasphere.co.uk/applications/water/pressure-flow-monitoring/

94. Flexim. (2020). Underground Pipeline Monitoring and Leak Detection. Flexim: When Measuring Matters. https://www.flexim.com/us/industry-solutions/water-industry/network-metering/underground-pipeline-monitoring-and-leak-detection

95. Pure Tech Ltd. (2020). Sahara leak detection tool used for speed, accuracy, and on-the-spot results. Pure Technologies: A Xylem Brand. https://puretechltd.com/articles/sahara-tool-for-speed-accuracy-and-on-the-spot-results/

96. About vGIS. (2020). VGIS. https://www.vgis.io/ar-mr-vr-gis-vgis-team-about-us/

97. Asian Development Bank. (2020). Using Artificial Intelligence for Smart Water Management Systems.

98. CityTaps. (2020). For Water Utilities - CityTaps. https://www.citytaps.org/for-water-utilities

99. Schneider Electric Global. (2021). EcoStruxure™ Geo SCADA Expert.

100. CGI. (2021). MOSAIC Supervisory Control and Data Acquisition (SCADA) system. https://www.cgi.com/en/solutions/mosaic

101. OSIsoft. (2021). https://www.osisoft.com/

102. Metasphere. (2021). Palette Web Hosted Telemetry Software. https://www.metasphere.co.uk/palette/

103. Leakmited. (2020). LEAKMITED: Water infrastructure management using satellite images and AI.

104. WHO, & UNICEF. (2017). Progress on drinking-water, sanitation and hygiene: 2017 update and SDG baselines.

105. FSM4. (2017). FSM4 Case Studies.

106. Interview with Richard Cheruiyot (Inspectorate Services Manager) and Brenda Anzagi (Information Communications and Technology Manager), WASREB. (2020).

107. Interview with Nienke Andriessen (Project Officer) and BJ Ward (Doctoral Researcher), Eawag, Fecal Sludge Management group. (2020).

108. Cummings, C., Langdown, I., Hart, T., Lubuva, J., & Kisela, H. (2016). What Drives Reform? Making Sanitation a Political Priority in Secondary Cities.

109. MajiData. (2011). Majidata - The Database for the Kenyan Water Sector. http://majidata.go.ke

110. Furlong, C. (2016). SFD (Excreta Flow Diagram) Promotion Initiative: Final Report.

111. WSUP. (2017). Research call: Urban Sanitation Costs and Low-Income Consumers' Willingness-to-Pay.

112. Peletz, R., Feng, A., MacLeod, C., Vernon, D., Wang, T., Kones, J., Delaire, C., Haji, S., & Khush, R. (2020). Expanding safe fecal sludge management in Kisumu, Kenya: a comparison of incentivized latrine pit emptying service model interventions. Journal of Water, Sanitation and Hygiene for Development, 10(4), 744–755. https://doi.org/10.2166/WASHDEV.2020.060

113. Focus Group (Port Harcourt, Nigeria). (2017).

114. Manual pit emptier, Port Harcourt, Nigeria. (2017).

115. Interview with Elizabeth Tilley, Professor, Malawi Polytechnic University. (2020).

116. Interview with Zaituni Kanenje, Pro-Poor Manager, NAWASSCO, Kenya. (2020).

117. Kisker, J., Renouf, R., & Drabble, S. (2019). Integrating mobile tech into sanitation services: Insights from Pula.

118. Sharma, A., & Bauer, G. (2019). Mobile for Development Utilities Annual Report Intelligent Utilities for All.

119. ONAS. (2014). Fecal sludge management program: Lessons learned. Quarterly Magazine of the ONAS FSM Program.

120. Ade, S. M., Musa, S. I., & Raji, M. A. (2014). Network Design and Tracking for Sewage Disposal and Management Using Geographic Information System and Remote Sensing For Mubi, Adamawa State, Nigeria. IOSR Journal of Environmental Science, Toxicology and Food Technology, 8(1), 66–73. https://doi.org/10.9790/2402-08146673

121. Hossain, M. D., Ninsawat, S., Sharma, S., Koottatep, T., & Sarathai, Y. (2016). GIS oriented service optimization for fecal sludge collection. Spatial Information Research, 24(3), 235–243. https://doi.org/10.1007/s41324-016-0024-z

122. Narendra, K. G., Swamy, C., & Nagadarshini, K. N. (2014). Efficient Garbage Disposal Management in Metropolitan Cities Using VANETs. Journal of Clean Energy Technologies, 2(3), 258–262. https://doi.org/10.7763/jocet.2014.v2.136

123. Daudey, L. (2018). The cost of urban sanitation solutions: A literature review. In Journal of Water Sanitation and Hygiene for Development (Vol. 8, Issue 2, pp. 176–195). IWA Publishing. https://doi.org/10.2166/washdev.2017.058

124. Tilley, E., Ulrich, L., Lüthi, C., Reymond, P., & Zurbrügg, C. (2014). Compendium of Sanitation Systems and Technologies (2nd Ed.).

125. Hutton, G., & Varughese, M. (2016). Technical Paper: the Costs of Meeting the 2030 Sustainable Development Goal Targets on Drinking Water, Sanitation, and Hygiene.

126. Haller, L., Hutton, G., & Bartram, J. (2007). Estimating the costs and health benefits of water and sanitation improvements at global level. Journal of Water and Health, 5(4), 467–480. https://doi.org/10.2166/wh.2007.008

127. Hutton, G., Haller, L., & Bartram, J. (2007). Global cost-benefit analysis of water supply and sanitation interventions. Journal of Water and Health, 5(4), 481–502. https://doi.org/10.2166/wh.2007.009

128. Whittington, D., Jeuland, M., Barker, K., & Yuen, Y. (2012). Setting Priorities, Targeting Subsidies among Water, Sanitation, and Preventive Health Interventions in Developing Countries. World Development, 40(8), 1546–1568. https://doi.org/10.1016/j.worlddev.2012.03.004

129. WSUP. (2018). Comparing the costs of different urban sanitation solutions in developing cities in Africa and Asia (Policy Brief).

130. Hutton, G. (2015, January). Why Choosing the Preferred Sanitation Solution Should Be More Like Grocery Shopping. World Bank Blog. https://blogs.worldbank.org/water/why-choosing-preferred-sanitation-solution-should-be-more-grocery-shopping

131. Crocker, J., Saywell, D., Shields, K. F., Kolsky, P., & Bartram, J. (2017). The true costs of participatory sanitation: Evidence from community-led total sanitation studies in Ghana and Ethiopia. Science of the Total Environment, 601–602, 1075–1083. https://doi.org/10.1016/j.scitotenv.2017.05.279

132. Trémolet, S., Kolsky, P., & Perez, E. A. (2010). Financing On-Site Sanitation for the Poor: a Six Country Comparative Review and Analysis.

133. Manga, M., Bartram, J., & Evans, B. E. (2020). Economic cost analysis of low-cost sanitation technology options in informal settlement areas (case study: Soweto, Johannesburg). International Journal of Hygiene and Environmental Health, 223(1), 289–298. https://doi.org/10.1016/j.ijheh.2019.06.012

134. Von Münch, E., & Mayumbelo, K. M. K. (2007). Methodology to compare costs of sanitation options for low-income peri-urban areas in Lusaka, Zambia. Water SA, 33(5), 593–602. https://doi.org/10.4314/wsa.v33i5.184017

135. Ulrich, L., Salian, P., Saul, C., Justrich, S., & Luthi, C. (2016). Assessing the Costs of on-Site Sanitation Facilities: Study Report.

136. Langergraber, G. (2014). CLARA Simplified Planning Tool. Sustainable Sanitation and Water Management Toolbox. https://sswm.info/planning-and-programming/decision-making/situation-and-problem-analysis/clara-simplified-planning-tool

137. Community Systems Foundation. (2015). Economics of Sanitation Toolkit. World Bank Group Global. https://www.communitysystemsfoundation.org/uploads/1/9/9/2/19920247/[p0424]_world_bank_india_esi_toolkit_enhancement_r2.pdf

138. Lüthi, C., Morel, A., Tilley, E., & Ulrich, L. (2011). Community-Led Urban Environmental Sanitation Planning (CLUES).

139. Bill & Melinda Gates Foundation. (2021). CWIS SAP: Citywide Inclusive Sanitation Services Assessment and Planning.

140. WASH Research Group School of Civil Engineering University of Leeds. (2020). CACTU$ Costing: Costs of Citywide Sanitation. http://cactuscosting.com/the-project/

141. SWA. (2020). WASH SDG Costing Tool. https://www.sanitationandwaterforall.org/tools-portal/tool/sdg-costing-tool

142. Veenkant, M., & Fonseca, C. (2019). Collecting life-cycle cost data for WASH services: A guide for practitioners. January.

143. World Bank. (n.d.). CWIS Planning Tool 1_4 (SERVER). http://200.58.79.50/fmi/webd/CWIS Planning Tool 1_4

144. Troeger, C., Forouzanfar, M., Rao, P. C., Khalil, I., Brown, A., Reiner, R. C., Fullman, N., Thompson, R. L., Abajobir, A., Ahmed, M., Alemayohu, M. A., Alvis-Guzman, N., Amare, A. T., Antonio, C. A., Asayesh, H., Avokpaho, E., Awasthi, A., Bacha, U., Barac, A., … Mokdad, A. H. (2017). Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet Infectious Diseases, 17(9), 909–948. https://doi.org/10.1016/S1473-3099(17)30276-1

145. Mboussou, F., Ndumbi, P., Ngom, R., Kassamali, Z., Ogundiran, O., Beek, J. Van, Williams, G., Okot, C., Hamblion, E. L., & Impouma, B. (2019). Infectious disease outbreaks in the African region: overview of events reported to the World Health Organization in 2018. Epidemiology and Infection, 147(e299), 1–8. https://doi.org/10.1017/S0950268819002061

146. Istepanian, R. S. H., & Al-Anzi, T. (2018). m-Health 2.0: New perspectives on mobile health, machine learning and big data analytics. Methods, 151(May), 34–40. https://doi.org/10.1016/j.ymeth.2018.05.015

147. Lleo, M. M., Lafaye, M., & Guell, A. (2008). Application of space technologies to the surveillance and modelling of waterborne diseases. Current Opinion in Biotechnology, 19(3), 307–312. https://doi.org/10.1016/j.copbio.2008.04.001

148. Alpher, J. (2016). Big Data and Analytics for Infectious Disease Research, Operations, and Policy. The National Academies Press, January. https://doi.org/10.17226/23654

149. WHO/WMO. (2016). Climate Services for Health: Improving public health decision-making in a new climate.

150. USAID. (2019). Artificial Intelligence in Global Health. In: From CII's Innovating for Impact Series.

151. George, D. B., Taylor, W., Shaman, J., Rivers, C., Paul, B., O'Toole, T., Johansson, M. A., Hirschman, L., Biggerstaff, M., Asher, J., & Reich, N. G. (2019). Technology to advance infectious disease forecasting for outbreak management. Nature Communications, 10(1), 8–11. https://doi.org/10.1038/s41467-019-11901-7

152. Polonsky, J. A., Baidjoe, A., Kamvar, Z. N., Cori, A., Durski, K., John Edmunds, W., Eggo, R. M., Funk, S., Kaiser, L., Keating, P., Le Polain De Waroux, O., Marks, M., Moraga, P., Morgan, O., Nouvellet, P., Ratnayake, R., Roberts, C. H., Whitworth, J., & Jombart, T. (2019). Outbreak analytics: A developing data science for informing the response to emerging pathogens. Philosophical Transactions of the Royal Society B: Biological Sciences, 374(1776). https://doi.org/10.1098/rstb.2018.0276

153. Pasetto, D., Finger, F., Camacho, A., Grandesso, F., Cohuet, S., Lemaitre, J. C., Azman, A. S., Luquero, F. J., Bertuzzo, E., & Rinaldo, A. (2018). Near real-time forecasting for cholera decision making in Haiti after Hurricane Matthew. PLoS Computational Biology, 14(5), 1–22. https://doi.org/10.1371/journal.pcbi.1006127

154. Hasan, M. A., Mouw, C., Jutla, A., & Akanda, A. S. (2018). Quantification of Rotavirus Diarrheal Risk Due to Hydroclimatic Extremes Over South Asia: Prospects of Satellite-Based Observations in Detecting Outbreaks. GeoHealth, 2(2), 70–86. https://doi.org/10.1002/2017gh000101

155. Chao, D. L., Roose, A., Roh, M., Kotloff, K. L., & Proctor, J. L. (2019). The seasonality of diarrheal pathogens: A retrospective study of seven sites over three years. PLoS Neglected Tropical Diseases, 13(8), 1–20. https://doi.org/10.1371/journal.pntd.0007211

156. Centers for Disease Control and Prevention. (2021). COVID-19 Mathematical Modeling. https://www.cdc.gov/coronavirus/2019-ncov/science/forecasting/mathematical-modeling.html

157. UNICEF, & WHO. (2019). Progress on household drinking water, sanitation and hygiene 2000-2017: Special focus on inequalities.

158. Hoque, S. F., & Hope, R. (2019). Examining the Economics of Affordability through Water Diaries in Coastal Bangladesh. Water Economics and Policy, 1950011. https://doi.org/10.1142/S2382624X19500115

159. Cook, P., Whittington, D., & Kimuyu, J. (2016). The costs of coping with poor water supply in rural Kenya. Water Resources Research, 52, 841–859. https://doi.org/10.1111/j.1752-1688.1969.tb04897.x

160. Hope, R., Thomson, P., Koehler, J., & Foster, T. (2020). Rethinking the economics of rural water in Africa. Oxford Review of Economic Policy, 36(1), 171–190. https://doi.org/10.1093/oxrep/grz036

161. Mitlin, D., Beard, V. A., Satterthwaite, D., & Du, J. (2019). Unaffordable and Undrinkable: Rethinking Urban Water Access in the Global South.

162. Satterthwaite, D., Beard, V. A., Mitlin, D., & Du, J. (2019). Untreated and Unsafe: Solving the Urban Sanitation Crisis in the Global South.

163. National Research Council. (2001). Under the Weather: Climate, Ecosystems, and Infectious Disease.

164. WHO. (2004). Using Climate to Predict Infectious Disease Outbreaks: A Review.

165. Interview with Mayank Midha, Director, GARV Toilets. (2020).

166. WHO. (2016). Quantitative microbial risk assessment: application for water safety management.

167. Haas, C. (2002). Progress and data gaps in quantitative microbial risk assessment. Water Science Technology, 46(11–12), 277–284.

168. Moran, K. R., Fairchild, G., Generous, N., Hickmann, K., Osthus, D., Priedhorsky, R., Hyman, J., & Del Valle, S. Y. (2016). Epidemic Forecasting is Messier Than Weather Forecasting: The Role of Human Behavior and Internet Data Streams in Epidemic Forecast. Journal of Infectious Diseases, 214. https://doi.org/10.1093/infdis/jiw375

169. WRBU. (2020). VectorMap: A web-based platform providing access and visualization of global vector distribution data. VectorMap. http://www.wrbu.org/docs/factsheet_vectormap_final.pdf

170. Universities Space Research Association (USRA). (2020). Global Monitoring and Mapping of Chikungunya Risk (CHIKRisk). https://vbd.usra.edu

171. Merkord, C. L., Liu, Y., Mihretie, A., Gebrehiwot, T., Awoke, W., Bayabil, E., Henebry, G. M., Kassa, G. T., Lake, M., & Wimberly, M. C. (2017). Integrating malaria surveillance with climate data for outbreak detection and forecasting: The EPIDEMIA system. Malaria Journal, 16(1), 1–15. https://doi.org/10.1186/s12936-017-1735-x

172. Sundram, B. M., Baha Raja, D., Mydin, F., Yee, T. C., Raj, K., & Kamaludin, F. (2019). Utilizing Artificial Intelligence as a Dengue Surveillance and Prediction Tool. J Appl Bioinforma Comput Biol, 8, 1. https://doi.org/10.4172/2329-9533.1000165

173. United Nations Educational Scientific and Cultural Organization (UNESCO), & Michigan State University (MSU). (2020). Global Water Pathogen Project. GWPP. http://www.waterpathogens.org

174. London Applied & Spatial Epidemiology Research Group (LASER), L. S. of H. and T. M. (2020). Global Atlas of Helminth Infections. GAHI. http://www.thiswormyworld.org

175. European Union. (2020). HEALTHY FUTURES. Healthy Futures Project. http://www.healthyfutures.eu

176. University of Maryland School of Medicine. (2021). Global Enteric Multicenter Study (GEMS). https://www.cbcb.umd.edu/datasets/gems-study-diarrheal-disease

177. Global Infectious Diseases and Epidemiology Network (GIDEON). (2021). https://www.gideononline.com/

178. Kotchi, S. O., Bouchard, C., Ludwig, A., Rees, E. E., & Brazeau, S. (2019). Using Earth observation images to inform risk assessment and mapping of climate change-related infectious diseases. Canada Communicable Disease Report, 45(5), 133–142. https://doi.org/10.14745/ccdr.v45i05a04

179. USAID. (2020). Famine Early Warning System Network (FEWS NET). FEWS NET. https://fews.net

180. Gething, P., Tatem, A., Bird, T., & Burgert-Brucker, C. R. (2015). DHS Spatial Analysis Reports No. 11: Creating Spatial Interpolation Surfaces with DHS Data.

181. Burgert-Brucker, C. R., Dontamsetti, T., Marshall, A. M. J., & Gething, P. W. (2016). DHS Spatial Analysis Reports No. 14: Guidance for Use of The DHS Program Modeled Map Surfaces.

182. Carrat, F., & Valleron, A.-J. (1992). Epidemiologic Mapping using the “Kriging” Method: Application to an Influenza-like Epidemic in France. American Journal of Epidemiology, 135(11), 1293–1300. https://doi.org/10.1093/oxfordjournals.aje.a116236

183. Guanais, F. (2013). Municipal-level Covariates of Health Status in Brazil: A Proposed Method for Data Interpolation. Revista Panamericana de Salud Pública, 34(3), 190–197.

184. WorldPop. (2020). Open Spatial Demographic Data and Research. WorldPop. http://www.worldpop.org

185. Interview with Alberto Wilde, Country Director, Global Communities, Ghana. (2020).

186. USAID. (2018). An Examination of CLTS’s Contributions Toward Universal Sanitation.

187. WaterAid, UNICEF, & International, P. (2017). Rethinking Rural Sanitation Approaches.

188. Interview with Sara Marks, Professor, Eawag. (2020).

189. Andrés, T. L. A., Borja-Vega, C., & Germán Sturzenegger, E. (2018). Innovations in WASH Impact Measures: Water and Sanitation Measurement Technologies and Practices to Inform the Sustainable Development Goals. In Directions in Development - Infrastructure. The World Bank. https://doi.org/doi:10.1596/978-1-4648-1197-5

190. Stuart, K., Peletz, R., Albert, J., Khush, R., & Delaire, C. (2021). Where does CLTS work best? Quantifying determinants of CLTS performance with evidence from four countries. Environmental Science & Technology, 55(6), 4064–4076. https://doi.org/10.1021/acs.est.0c05733

191. UN-Water. (2014). Coordination of Water Actions at the Country Level: A Report of the UN-Water Task Force on Country Level Coordination.

192. Jia, P., Anderson, J. D., Leitner, M., & Rheingans, R. (2016). High-resolution spatial distribution and estimation of access to improved sanitation in Kenya. PLoS ONE, 11(7), 1–21. https://doi.org/10.1371/journal.pone.0158490

193. Wong, K. L. M., Brady, O. J., Campbell, O. M. R., & Benova, L. (2018). Comparison of spatial interpolation methods to create high-resolution poverty maps for low-and middle-income countries. J R Soc Interface, 15(147). https://doi.org/10.1098/rsif.2018.0252

194. SIASAR Global: Rural Water and Sanitation Information System. (2017). https://globalsiasar.org/en

195. Global Water Challenge. (2020). Water Point Data Exchange (WPdx). https://www.waterpointdata.org/

196. Kar, K., & Chambers, R. (2008). Handbook on Community-Led Total Sanitation (Vol. 44). Plan International.

197. Interview with Bal Mukunda Kunwar, Business Development Officer, Helvetas Swiss Intercooperation Nepal. (2020).

198. Interview with Aliocha Salagnac, Information Management Systems Specialist, UNICEF. (2020).

199. Interview with Rob Bain, Statistics and Monitoring Specialist, UNICEF. (2020).

200. Interview with Christopher Kanyesigye, Research and Development Manager, NWSC Uganda. (2020).

201. Interview with Aaron Salzberg, Director, UNC Water Institute. (2020).

202. Interview with Mbaye Mbeguere, Senior WASH Manager and Urban Focal Point, WaterAid Senegal. (2020).

203. Interview with Matteus Van der Velden, Manager, Asia Regional Unit, WSSCC. (2020).

204. Interview with Susanna Smets (Senior Regional Water Supply and Sanitation Specialist) and Kristofer Welsien (Senior Water Supply and Sanitation Specialist), World Bank, Global Water Practice. (2020).

205. ITech Mission. (n.d.). SDGs Global Dashboard. Explore, Monitor and Visualize SDGs Data. https://www.sdgsdashboard.org/

206. WHO/UNICEF. (n.d.). JMP Data. https://washdata.org/data

207. AtlasAI. (2021). Helping You Decide Where To Invest In Africa and South Asia.

208. Wiesmann, U., Kiteme, B., & Mwangi, Z. (2016). Socio-economic atlas of Kenya: depicting the national population census by county and sub-location (2nd ed.). Kenya National Bureau of Statistics, Nairobi. Centre for Training and Integrated Research in ASAL Development, Nanyuki. Centre for Development and Environment, Bern. https://doi.org/10.7892/boris.83693

209. Shapiro, J., Tribbe, J., Stuart, K., Mok, S., & Helme, N. (2021). SanPlan: The Sanitation Planning Tool. In USAID Webinar.

210. WHO, & UNICEF. (2017). Progress on household drinking water, sanitation and hygiene 2000-2017. Special focus on inequalities. New York: United Nations Children's Fund (UNICEF) and World Health Organization, 2019.

211. Interview with Issifu Adama, WASH Officer, UNICEF Ghana. (2020).

212. Andres, L., Thibert, M., Lombana, C., Danilenko, A., Joseph, G., & Borja-Vega, C. (2019). Doing More with Less: Smarter Subsidies for Water Supply and Sanitation.

213. Abramovsky, L., Andrés, L., Joseph, G., Pablo-Rud, J., Sember, G., & Thibert, M. (2020). Study of the Distributional Performance of Piped Water Consumption Subsidies in 10 Developing Countries (Working Paper).

214. Cardenas, H., & Whittington, D. (2019). Magnitude and Distribution of Electricity and Water Subsidies for Households in Addis Ababa, Ethiopia. https://doi.org/10.1596/1813-9450-9025

215. Fuente, D., Gakii Gatua, J., Ikiara, M., Kabubo-Mariara, J., Mwaura, M., & Whittington, D. (2016). Water and sanitation service delivery, pricing, and the poor: An empirical estimate of subsidy incidence in Nairobi, Kenya. Water Resources Research, 52, 4845–4862. https://doi.org/10.1002/2015WR018375

216. Aquaya Institute, (2019). Research on the Impact of Targeted Subsidies Within Open Defecation Free (ODF) Communities.

217. Alkire, S., & Seth, S. (2013). Selecting a Targeting Method to Identify BPL Households in India. Social Indicators Research, 112(2), 417–446. https://doi.org/10.1007/s11205-013-0254-6

218. Aryeetey, G., Jehu-Appiah, C., Spaan, E., D'Exelle, B., Agyepong, I., & Baltussen, R. (2010). Identification of poor households for premium exemptions in Ghana's National Health Insurance Scheme: empirical analysis of three strategies. Tropical Medicine & International Health, 15(12), 1544–1552.

219. Dershem, L., Saidulloev, F., Nadareishvili, M., Arnold, C., & Rittmann, J. (2013). Using a Proxy Means Test for Targeting in a Conditional Cash Transfer Program.

220. Water.org. (2020). Water Credit Initiative. https://water.org/solutions/watercredit/

221. Tarafder, K. H., & Custodio, B. B. (2019). ASA Philippines' Water and Sanitation Financing Program: Leveraging a Quality of Life Financing to Obtain Financial Gain. OIDA International Journal of Sustainable Development, 12(8), 21–46.

222. IDinsight. (2013). Microfinance loans to increase sanitary latrine sales.

223. Interview with Chris Nicoletti (Senior Director of Impact and Analytics) and Abdul-Mumin Damba Tahidu (Country Manager), iDE Ghana. (2020).

224. Kshirsagar, V., Wieczorek, J., Ramanathan, S., & Wells, R. (2017). Household poverty classification in data-scarce environments: a machine learning approach. 31st Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA.

225. Poulin, C., Trimmer, J., Press-Williams, J., Yachori, B., Khush, R., Peletz, R., & Delaire, C. (n.d.). Performance of a novel AI-based proxy means test in comparison to other methods for targeting pro-poor water subsidies in Ghana (Submitted). Development Engineering.

226. Dupriez, O., Solatorio, A., & Bank, T. W. (n.d.). Machine Learning for Dynamic Survey.

227. USAID Indonesia Urban Water Sanitation and Hygiene Penyehatan Lingkungan untuk Semua (IUWASH PLUS). (2019). Tracking impact for poor and vulnerable households.

228. Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., & Burke, M. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. Nature Communications, 11(1), 1–11. https://doi.org/10.1038/s41467-020-16185-w

229. Facebook Data for Good. (2021). Relative Wealth Index.
https://dataforgood.facebook.com/dfg/tools/relative-wealth-index

230. Watmough, G. R., Marcinko, C. L. J., Sullivan, C., Tschirhart, K., Mutuo, P. K., Palm, C. A., & Svenning, J. C. (2019). Socioecologically informed use of remote sensing data to predict rural household poverty. Proceedings of the National Academy of Sciences of the United States of America, 116(4), 1213–1218. https://doi.org/10.1073/pnas.1812969116

231. Logar, T., Bullock, J., Nemni, E., Bromley, L., Quinn, J., & Luengo-Oroz, M. (2020). PulseSatellite: A Tool Using Human-AI Feedback Loops for Satellite Image Analysis in Humanitarian Contexts. Proceedings of the AAAI Conference on Artificial Intelligence, 34(9), 13628–13629. https://doi.org/10.1609/aaai.v34i09.7101

232. Ravallion, M. (2018). Should the Randomistas Continue to Rule?

233. York, P., & Bamberger, M. (2020). Measuring results and impact in the age of big data: The nexus of evaluation, analytics, and digital technology. March.

234. Venkataramanan, V., Crocker, J., Karon, A., & Bartram, J. (2018). Community-Led Total Sanitation: A Mixed-Methods Systematic Review of Evidence and Its Quality. Environmental Health Perspectives, 126(2), 17. https://doi.org/10.1289/EHP1965

235. Zuin, V., Delaire, C., Peletz, R., Cock-Esteb, A., Khush, R., & Albert, J. (2019). Policy Diffusion in the Rural Sanitation Sector: Lessons from Community-Led Total Sanitation (CLTS). World Development, 124, 104643. https://doi.org/10.1016/j.worlddev.2019.104643

236. Interview with Antoinette Kome, Global Sector Coordinator WASH, SNV. (2020).

237. UN Global Pulse. (2014). Analyzing Social Media Conversations To Understand Public Perceptions of Sanitation (No. 5; Global Pulse Project Series).

238. Hilton Foundation. (2020). Conrad N . Hilton Foundation Safe Water Initiative: Strategy, Measurement, Evaluation and Learning: Selecting Comparator Districts.

239. Stanford Program on Water Health and Development. (2020, April 3). Using data to inform philanthropy decision-making.

240. United Nations Office for the Coordination of Humanitarian Affairs (OCHA). (2019). Data Responsibility Guidelines, Working Draft.

241. The Earth Genome. (2021). https://www.earthgenome.org/

242. World Resources Institute. (2021). Data. https://www.wri.org/

243. Water Peace and Security. (2021). Global Tool. https://waterpeacesecurity.org/map

244. SERVIR GLOBAL | Connecting Space to Village. (2021). https://www.servirglobal.net/

245. OpenET. (2021). https://openetdata.org/

246. Duke University. (2021). Internet of Water: Connecting Water Data to Improve Outcomes. https://internetofwater.org/

247. California Department of Water Resources. (2021). Flood-Managed Aquifer Recharge (Flood-MAR). https://water.ca.gov/programs/all-programs/flood-mar

248. Texas Water Development Board. (2021). Water Data For Texas. https://waterdatafortexas.org/

Appendix 1

Examples of existing large-scale WASH data initiatives

Earth Genome

Earth Genome, a non-profit startup, applies big environmental data to offer solutions and insights expressed in financial terms to corporations or government agencies.[241] They develop visualization tools and applications to support decision makers in evaluating the financial, environmental, and social impacts of investments and interventions. Earth Genome addresses water scarcity with the Green Infrastructure Tool and aquifer recharge with the Groundwater Recharge Assessment Tool. They offer custom applications for source water protection, watershed resilience, and watershed health.

World Resources Institute

The World Resources Institute, a non-profit think tank, responds to seven urgent global challenges: climate, energy, food, water, forests, sustainable cities, and oceans. They aim to reduce poverty, grow economies, and protect natural ecosystems. In designing solutions to address these challenges, the World Resources Institute produces reports, peer-reviewed publications, charts, maps, datasets, and visual resources. They aim to turn information into action through funded projects including data platforms such as Water Peace and Security, Energy Access Explorer, Resource Watch, and Climate Watch.[242]

Global Water Intelligence

Global Water Intelligence (GWI), a for-profit company, provides business information and networking opportunities for the water industry in multiple countries.[75] They provide users with reports on the technology and innovation landscape in the water sector, in addition to a project tracker to filter global water projects by type (build, operate, finance), categories (desalination, utility, wastewater treatment plant), location, capacity, and expected cost. GWI connects actors by offering interest groups tailored to water utilities, water reuse leaders, and industries such as oil and gas. They manage a platform for research and knowledge exchange and networking. In addition, GWI organizes events such as the Global Water Summit and the Ultrapure Micro conference.

GWI maintains an online platform that displays data on the water market by country or by industry, including information on market size, growth rate, business constraints, and opportunities. They offer market research reports that provide insights in the international water and desalination water sectors; these reports include forecasts and downloadable data.

Water, Peace and Security (WPS)

The WPS partnership is a collaboration between governments, academic institutions, and CSOs, including the Netherlands Ministry of Foreign Affairs, IHE Delft Institute for Water Education, World Resource Institute, Deltares (a Dutch independent research institute), Hague Centre for Strategic Studies, Wetlands International, and International Alert. They provide interactive maps using water and socioeconomic datasets.[243]

SERVIR

SERVIR is a joint initiative of the US National Aeronautics and Space Administration (NASA), USAID, and regional organizations in Africa, South Asia, South and Central America, to help low- and middle-income countries use and access information provided by Earth-observing satellites and geospatial technologies.[244] SERVIR provides data, maps, and visualization tools to better manage climate risks, natural disasters, food security, water resources, and land use change. Users can directly download data and images from the website. SERVIR also offers training events through webinars, classes, and field training.

India Water Tool

The India Water Tool helps companies and users identify water risks within India.[74] The World Business Council for Sustainable Development coordinated the project, and the Confederation of Indian Industry-Triveni Water Institute and World Resources Institute provided technical expertise. The website provides point data with groundwater level and surface water quality, groundwater information (e.g., exploitation of aquifers, type of aquifers), and water stress indicators (e.g., rainfall, surface water availability, water deficit).

Environmental Defense Fund (United States)

The Environmental Defense Fund (EDF) is an American non-profit organization that works on advocacy campaigns and political lobbying to safeguard the environment. EDF is leveraging new technologies such as sensors and satellites to fill data gaps: “Sensors, combined with powerful data analytics are transforming environmental protection,” according to the President of EDF. For example, EDF developed sensors to measure hyperlocal air pollution, to detect pollution from oil and natural sources, and to collect data on chemical exposure. In partnership with NASA, the Desert Research Institute, Google, and others, their OpenET (evapotranspiration) platform provides water balance data to farmers, water managers, and the public.[245]

Internet of Water (United States)

Internet of Water (IoW) began as a collaboration between the Aspen Institute and Duke University to improve water data infrastructure in the United States.[246] They aim to accelerate the development of open data and provide up-to-date data for real-time water management. To do so, they built a network that connects data producers (e.g., water utilities), data hubs (e.g., United States Geological Survey, National Groundwater Monitoring Network Portal), and data users together to facilitate sharing and integration of water data. IoW has piloted projects in different states in the United States (New Mexico, California, Texas), where they support stakeholders in using integrated water data for sustainable water management. IoW resources consist of an inventory of the types of water data made available by public agencies, a library for finding reports, educational materials, and a page to help users to understand data definitions.

California's Open and Transparent Water Data Act

To improve California's water management in the face of climate change and meet environmental and human needs, the 2016 Open and Transparent Water Data Act aimed to provide the state with complete, accessible, and usable data.[6] This led to the creation of the California Department of Water Resources Open Data Portal, which has 253 datasets (as of 2021), covering agriculture, land use, weather and climate, and other information that might be relevant for water policy decision-making. Users can share and publish datasets through the site. During platform development, a needs assessment revealed that state and federal agencies provided 90% of the data.[6] The remaining 10% came from research institutions and non-profit organizations. The study also identified some data gaps on economic information (water price, willingness to pay, and economic impact), on characterization of water rights and on information on groundwater pumping. The California Department of Water Resources (DWR) also operates a Water Data Library, which publishes government datasets on groundwater.

California Department of Water Resources – Flood-MAR

The Flood-Managed Aquifer Recharge (Flood-MAR) is a program developed by the California DWR for improving aquifer recharge.[247] They designed this strategy with a research advisory committee consisting of roughly 200 subject-matter experts. The committee developed a research and data plan to prioritize needed research and tools and ensure the program is data driven. They provide publications and reports to help stakeholders implement the Flood-MAR technique.

Texas Water Development Board – Water Data for Texas

The Water Data for Texas website synthesizes and communicates water data to scientists, policymakers, and the public.[248] Their collaborations with different institutions (National Oceanic and Atmospheric Administration, United States Department of Agriculture, National Drought Mitigation Center, University of Nebraska Lincoln, PRISM Climate Group at Oregon State University, and the United States Geological Survey) allow them to offer interactive maps and downloadable data on water and climate conditions. The site provides the status of water supply reservoirs, a weekly map of drought conditions, monthly anomaly precipitation and temperatures, monthly streamflow percentiles, and rainfall and evaporation forecasts. They also offer groundwater levels and water quality data in coastal areas.

Appendix 2

Decision-maker interviewees

Name	Title*	Organization*
Aaron Salzberg	Director	The Water Institute at the University of North Carolina, Chapel Hill (UNC)
Abdul-Mumin Damba Tahidu	Country Manager	iDE Ghana
Ada Oko-Williams	Senior Manager, Sanitation	WaterAid
Adam Harvey	Managing Director	Whave
Adrien Couton	Partner, Water Practice Director	Dalberg
Alberto Wilde	Country Director	Global Communities, Ghana
Aliocha Salagnac	Information Management Systems Specialist	United Nations Children's Fund (UNICEF)
Andrés Hueso	Senior Policy Analyst	WaterAid
Andy Robinson	Independent Consultant, Water Supply and Sanitation	
Angella Rinehold	Consultant	World Health Organization (WHO)
Antoinette Kome	Global Sector Coordinator for Water, Sanitation, and Hygiene (WASH)	Netherlands Development Organisation (SNV)
Ariane Schertenleib	Data Communications Specialist	Swiss Federal Institute of Aquatic Science and Technology (Eawag)
Aude-Sophie Rodella	Senior Economist	World Bank, Global Water Practice
Bal Mukunda Kunwar	Business Development Officer	Helvetas Nepal
BJ Ward	Doctoral Researcher	Eawag, Fecal Sludge Management Group
Brenda Anzagi-Sudi	Information Communications and Technology Manager	Water Services Regulatory Board (WASREB) Kenya
Carolien Van der Voorden	Head, Technical Support Unit	Water Supply and Sanitation Collaborative Council (WSSCC)

Name	Title*	Organization*
Chris Cormency	Chief of WASH	UNICEF Mozambique
Chris Nicoletti	Senior Director of Impact and Analytics	iDE
Christopher Print	Senior Land and Water Advisor	Food and Agriculture Organization of the United Nations (FAO)
Clarissa Brocklehurst	Independent Consultant, Water Supply and Sanitation	
Claudia Ruiz Vargas	Researcher	International Groundwater Resources Assessment Centre (IGRAC)
Christopher Kanyesigye	Research and Development Manager	National Water and Sewerage Corporation (NWSC) Uganda
Duncan McNicholl	Director	Uptime
Eduardo Perez	Technical Director, Global WASH and Health	Global Communities
Elizabeth Tilley	Professor	Malawi Polytechnic University
Emmanuel Opong	Country Program Manager	World Vision
Esha Zaveri	Economist	World Bank, Global Water Practice
Fiona Zakaria	Research Fellow	University of Leeds
François Bertone	Senior Water Resource Management Specialist, Groundwater	World Bank, Global Water Practice
Frédéric Bergeron	General Manager	Whave
George Wainaia	Doctoral Researcher	Eawag
Gerard Soppe	Senior Water and Sanitation Expert	World Bank
Greg Lestikow	Global WASH Director	iDE
Habab Taifour	Senior Water Resources Specialist	World Bank Ethiopia
Harrison Kwach	Independent Consultant, Urban Waste Management	
Heather Bischel	Assistant Professor	University of California, Davis
Innocent Tumwebaze	Country Director	Water Pathogen Knowledge to Practice Project (Water-K2P)
Isaac Kega	Monitoring and Evaluation Specialist	Kenya Water Sector Trust Fund

Name	Title*	Organization*
Issifu Adama	WASH Officer	UNICEF Ghana
Jamie Myers	Research and Learning Manager, Sanitation Learning Hub	Institute of Development Studies (IDS) UK
Jason Russ	Economist	World Bank, Global Water Practice
Jemima Sy	Program Manager, Public-Private Infrastructure Advisory Facility (PPIAF)	World Bank
Joseph Githinji	General Manager, Operations and Senior Manager, Customer Support	Sanergy Kenya
Kristoffer Welsien	Senior Water Supply & Sanitation Specialist	World Bank
Lindsey Noakes	Co-Founder	Gather.hub
Luis Andrés	Sector Leader and Lead Economist	World Bank
Matteus Van der Velden	Manager, Asia Regional Unit	WSSCC
Mayank Midha	Director	GARV Toilets India
Mbaye Mbeguere	Senior WASH Manager and Urban Focal Point	WaterAid Senegal
Michael Gnilo	Sanitation and Hygiene Specialist	UNICEF
Michael Kropac	Senior Partner	Seecon Switzerland
Neno Kukuric	Director	IGRAC
Nienke Andriessen	Project Officer	Eawag, Fecal Sludge Management Group
Patrick Moriarty	Chief Executive Officer	International Water and Sanitation Centre (IRC)
Philip Oyamo	Senior Project Manager	Water & Sanitation for the Urban Poor (WSUP) Kenya
Rafael Catalla	Programme Manager	Plan International
Ramon Brentführer	Project Manager, Groundwater Policy Advice	Federal Institute for Geosciences and Natural Resources (BGR) Germany
Richard Cheruiyot	Inspectorate Services Manager	WASREB Kenya
Rick Johnston	Technical Officer	WHO-JMP (Joint Monitoring Programme)

Name	Title*	Organization*
Rob Bain	Statistics and Monitoring Specialist	UNICEF
Robel Lambisso	Technical Program Manager	World Vision, Ghana
Roland Liemberger	Independent Consultant, Non-Revenue Water (NRW) Reduction	
Samuel Mwanangombe	Senior Global Advisor - Design, Monitoring, Evaluation and Research for WASH	World Vision
Sara Marks	Professor	Eawag
Stella Warue	Programme Officer	Kenya Water Sector Trust Fund
Susanna Smets	Senior Regional Water Supply and Sanitation Specialist	World Bank, Water Global Practice
Tariya Yusuf	Programme Support Advisor, Sanitation	WaterAid
Vincent Post	Hydrogeologist	BGR Germany
Zaituni Kenenje	Pro-Poor Manager	Nakuru Water and Sanitation Services Company (NAWASSCO) Kenya

* At time of interview (March–June 2020)

Appendix 3

Decision-maker interview questions

1) General questions [1]

a) Please describe the programs/projects that you are currently operating.
b) What is your vision for the next 5–10 years? What will be the big challenges?
c) Who are decision-makers? What decisions do you make regularly?
d) Are there any challenges or important decisions your organization faces on a routine basis for which you wish had more information?

2) Data-specific questions

a) Program management

i) How do the decision-makers use data and for which purpose? Who is using data for decision-making?
ii) How does data guide your institution in management of decision-making processes? What kind of data do you use to make decisions? Do you make any organizational decisions with that type of data?
iii) Do you have any reporting requirements to external stakeholders on program/project progress? What data are reported?

b) Program design and implementation

i) What type of data (any type) does your organization actively use on a daily basis? Does data influence program implementation?
ii) Do you think your decisions would be any different if you had more data? What are the challenges with data? What type of data that you don't have could help?
iii) For programs: Do you do any sort of monitoring and evaluation?

3) Data management and technical innovation questions

a) Data management

i) What is current staff capacity regarding data management and analysis? What are staff capacity strengths and gaps in relation to this?
ii) Do you want to develop a data team within your organization or do you prefer to work with consultants?
iii) How often are you reviewing data - Weekly? Monthly? Quarterly? What data are reported? In what format are they reported? Excel table? Graphs? Aggregate statistics? Digital or paper form? How do you integrate new data with existing data?
iv) Does your organization use any secondary data?
v) What type of data would you want available or easily accessible for your organization? For example, the data can be organizational, program-specific, sector-wide, etc.

[1] Part 1–3 questions used for all interviewees.

b) Technical innovations

i) What software platforms or tools do you currently use to view data? To analyze data? What resources or skill sets does the organization have for data analysis/software or IT maintenance? Basic/intermediate/advanced?

ii) Are you happy with the current technologies? Would you be open to new technologies? Which technological innovations can facilitate the evidence-based decision-making process and how?

4) Hypothesis-specific questions[2]

a) Water – general:

i) If more information was available about water and sanitation service levels in cities, how do you think this would affect the allocation of grants to counties and water service providers (WSPs)?

ii) If utility managers were able to forecast water consumption, to what extent would that allow them to plan and operate effectively, particularly with respect to urbanization and climate change?

iii) If cost models for subsidizing water during emergencies were available, how do you think it would have impacted the response of counties and WSPs to COVID-19?

b) Non-revenue water:

i) What innovations would you like to see implemented in water utilities to address the problem of NRW? To improve service coverage in low-income areas?

ii) If utilities had real-time data on flow rate and pressure in the piped network, how could that affect their ability to identify anomalies such as illegal connections or leaks?

iii) If utility managers had clear information on the leading causes of NRW (bill collection vs. l eaks etc.), to what extent would that affect their ability to address the issue?

iv) What do you see as the biggest research needs with respect to NRW reduction in Africa and Asia?

v) To what extent can the evidence base regarding technologies such as real-time sensors, prepaid meters and satellite-based leak detection be strengthened?

vi) To what extent can the evidence base regarding "softer" approaches to water loss reduction (e.g., incentives for customers and meter readers to report leaks and illegal connections) be reinforced?

vii) Are there pieces of evidence or knowledge that would help attract more finance towards NRW reduction?

c) Groundwater quality:

i) From your perspective, what are the biggest information gaps with respect to groundwater quality?

ii) What decisions would be easier if more groundwater quality information were available?

iii) Are you aware of specific countries who are particularly demanding of this information?

iv) How could high-resolution maps of groundwater salinity (or nitrates or fecal contamination) support the activities of the World Bank or its partners?

[2] Part 4 questions used if applicable to interviewee's organization or role.

d) Sanitation – general:

i) Given your current access to data, do you feel you have a good understanding of where there is improved sanitation versus unimproved sanitation?

ii) Are there tools you would like to have access to to help you understand where there is the greatest need for improving household sanitation infrastructure?

iii) What sources of data do you rely on for financial information as it relates to sanitation? Do you feel like this data source is accurate? Up to date? Reliable?

iv) What additional financial information about sanitation would you find useful in making decisions? What decisions would that financial information influence?

v) Do you think climate change will impact existing sanitation infrastructure? If so, how? What information would be helpful to understand how?

e) Sanitation – urban:

i) Given your current access to data, do you feel you have a good understanding of where there is improved versus unimproved sanitation servicing (i.e., emptying)?

ii) Do you believe open defecation in urban settings continues to be an issue? If so, do you think urban residents see it as an issue? What information would be helpful to share with residents to make them aware of the issue? If money weren't an issue, how would you address urban open defecation? Do you have any sense of what that kind of solution would cost?

f) Sanitation – rural:

i) Are you able to measure the success of your programming? If so, how do you measure success? Is there additional data that would help to measure success?

ii) Is it difficult to identify which communities to target for sanitation interventions? If so, what makes it difficult? What additional information would help you determine the best communities for interventions?

g) Cross-cutting/WASH – general:

i) What method(s) do you use to identify intervention areas? Do you have a sense of what prior interventions those areas may have been previously exposed to? If so, how were you able to access that data? If not, have you ever started an intervention to then realize a prior intervention had been implemented in that area? What information would be helpful to better target intervention areas?

h) Cross-cutting/WASH – subsidy targeting:

i) What are the biggest research needs with respect to subsidy targeting?

ii) To what extent could the development of rapid targeting tools help support the World Bank's efforts with respect to subsidy targeting?

iii) Are there locations (countries, cities) that are particularly opportune for the development of such tools, for example because local authorities are demanding support for targeting subsidies?

aquaya

CONRAD N. HILTON
FOUNDATION