aquaya

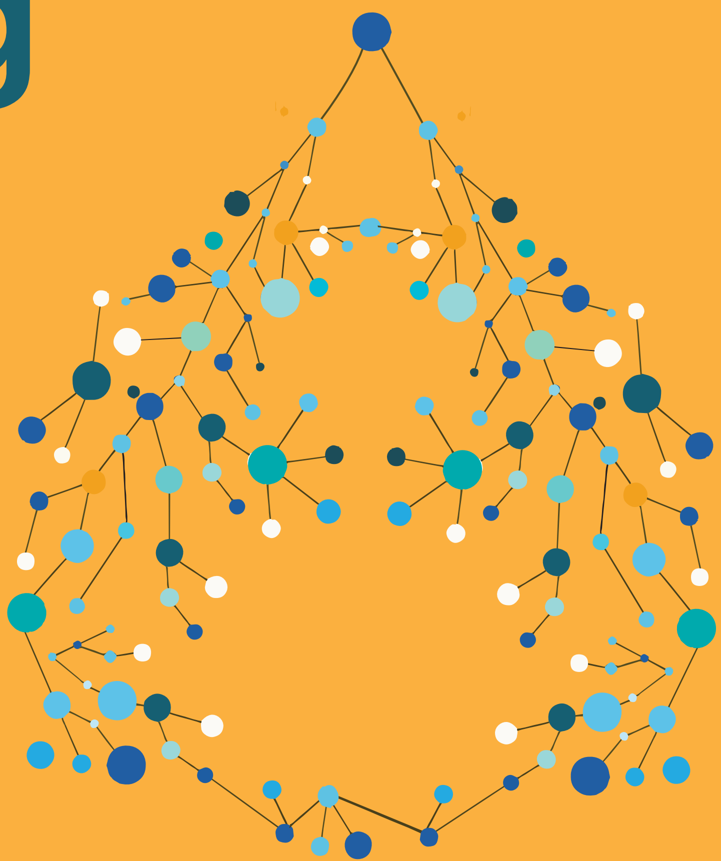
CONRAD N. HILTON FOUNDATION

January 2022

Data for Decision-Making

Water and Sanitation in Low-Resource Settings

Annex 1: Analysis of Results by Use Cases



Contents

Acronyms and Abbreviations

app	software application, especially as downloaded to a mobile device; may also signify a web-based or desktop computer application
BGR	German Federal Institute for Geosciences and Natural Resources
CACTUS	Climate and Costs in Urban Sanitation
CLTS	community-led total sanitation
CSO	civil society organization
DHS	Demographic and Health Surveys
DMA	district monitoring areas
FEWS	Famine Early Warning System
FSM	fecal sludge management
GIS	geographic information systems
GLAAS	UN-Water Global Analysis and Assessment of Sanitation and Drinking-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. 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
ICT	information and communications technology
IRC	International Rescue Committee
IWA	International Water Association
JMP	Joint Monitoring Programme
LIA	low-income area
MICS	Multiple Indicator Cluster Surveys
MFI	microfinance institution
NRW	non-revenue water
PBC	performance-based contract
PPIAF	Public-Private Infrastructure Advisory Facility

RCT	randomized controlled trial
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
UN	United Nations
UNICEF	United Nations Children's Fund
USAID	United States Agency for International Development
USD	United States Dollars
VANET	vehicular ad-hoc network
WASH	water, sanitation, and hygiene
WASREB	Kenyan Water Services Regulatory Board
WHO	World Health Organization
WPDx	Water Point Data Exchange
WSN	wireless sensor network
WSP	water service provider
WSSCC	Water Supply and Sanitation Collaborative Council
WSUP	Water & Sanitation for the Urban Poor

Approach

This document is an accompaniment to the summary report, Data for Decision-Making: Water and Sanitation in Low-Resource Settings. Sixty-seven decision-makers were invited to participate in the project, representing a broad cross-section of water, sanitation, and hygiene (WASH) organizations. Of the 10-person Aquaya research team, one or two researchers who were trained in qualitative methods and already familiar with the interviewee[s] administered questions during semi-structured phone or video calls between March and June 2020. Most decision-makers were interviewed individually, although some group interviews (focus groups) included up to four decision-makers representing the same organization. The full interview guide is available as Appendix 2 of the summary report. The questions included both a set of general questions for all interviewees and specific questions regarding predetermined use case hypotheses, tailored as applicable to the decision-makers' professional organizations or roles. Written notes were recorded during the interviews. 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 characteristics of each use case. The following categories were elicited:

Objective: Decision, goal, or desired action.

The objective highlights the stakeholders involved, how they typically meet their needs, and what types of decisions they make. Some objectives are simpler and only highlight one need for one decision; however, most use cases are more complex and describe multiple data needs or needs that affect multiple decisions.

Description: Key context and background information.

The description has basic information to put the data needs into context.

Decision Status Quo: How the decision-makers work in the absence of novel data science advances.

The decision status quo(s) demonstrates how decision-makers typically make key decisions related to the use cases.

Demand: Why the data need is critical for WASH decision-makers.

This section pulls from interviews with key sector stakeholders to demonstrate the need or desire for data-based solutions.

Other Data Applications: How the use case or its outputs could be used for other decisions.

The other data applications section addresses how else this solution might be applicable, for example, in other sectors or contexts.

Existing and Upcoming Innovations: Innovations directly relevant to WASH use cases (or potentially translatable).

This section highlights what is already happening to address the use case needs. Some innovations might be in a pilot or development phase, whereas others might be smoothly running software applications accessible from a computer or mobile device. Some examples might have drawbacks that could be improved upon, or address only a subcomponent of the use case. Innovation ordering does not imply any relative ranking. The Aquaya Institute ("Aquaya") has not independently reviewed the validity or performance of specific technologies or manufacturer claims described in this report; thus, the information is provided solely for reference. The examples provided are not exhaustive, as new organizations come onto the market or merge regularly and existing organizations continually upgrade their product and service offerings.

Participants: The main decision-makers who use the data and how they use it.

This section describes general decision-maker roles and existing levels of experience.

Outputs: Products or services that could meet decision-making needs.

Products or services of interest for further development by data providers such as Aquaya and other organizations, owing to the demand demonstrated by decision-makers.

Workflow: The progression of steps and specific actions data providers might take to accomplish use case objectives.

For use cases that Aquaya has already been involved in developing, the workflow is better understood and sometimes tested. More challenging or newer use cases may have less well-defined workflows.

Data sources: Existing data sources and data gaps.

For use cases Aquaya has already been involved in developing, the breadth and pros and cons of existing data sources are better understood. A separate Aquaya project, Project W, aims to exhaustively compile data sources usable by WASH decision-makers.

Barriers: Challenges, barriers, or uncertainties to achieving the use case objectives.

The barriers included rely on a broad sample of WASH decision-maker input, although execution of the use case might raise additional challenges.

Use Cases

1. WATER: FORECASTING GROUNDWATER QUALITY AND QUANTITY

OBJECTIVE

To provide governments (national, regional, or local), civil society organizations (CSOs), water suppliers/utilities, multilateral UN agencies, and the private sector (e.g., banks, consultants) with user-friendly models and maps (accessible through an application interface) that: i) predict groundwater quality and ii) predict groundwater quantity.

Decisions (groundwater quality):

- Governments would use the data to ensure access to safe water and protect public health through the development of effective regulations, factoring in population growth and economic development.
- In conjunction with climate models, governments would use the data to mitigate climate risks by identifying water risk areas and developing appropriate water management strategies.
- Governments would use applications to develop appropriate legal frameworks to protect vulnerable resources at different time scales.
- Governments, CSOs, and the private sector would use the data to predict the proportion of aquifers with good water quality, thus tracking or guiding resource protection efforts.
- Implementing organizations would use the data to prioritize programming in areas with degraded water quantity or quality.

Decisions (groundwater quantity):

Water suppliers/utilities would use applications to plan for infrastructure expansion and maintenance and to anticipate possible future water shortages.

- Implementing organizations and governments would use applications to develop plans for emergencies (e.g., droughts, flood), and support emergency response and recovery efforts.
- Implementing organizations would use applications to prioritize program implementation to focus on areas facing water shortage.

DESCRIPTION

Governments have committed to ensuring safe and affordable drinking water for all by 2030 (United Nations Sustainable Development Goal [SDG] 6.1). A safely managed drinking water service has to meet the following criteria: accessible on premises, available when needed, and free from contamination. In 2015, 31% of the global population (i.e., 2.2 billion people) still did not have access to safely managed drinking water services, and the majority of these 2.2 billion people were in Africa. Even though 181 countries had achieved more than 75% coverage of at least basic drinking water services, more than three quarters of sub-Saharan African countries had less than 75% coverage. The primary challenge in providing 100% coverage is providing access to safely managed drinking water services in rural areas and low-income urban areas. In addition, climate change (in combination with increased population density) threatens water resources in Africa, which will affect both the quality and the quantity of groundwater.[2,3]

Regarding water quality, SDG 6.3 aims to improve ambient water quality (including rivers, lakes, and groundwater) by reducing pollution from the point sources (e.g., wastewater and industries) and from non-point sources (e.g., runoff from urban and agricultural land) that may penetrate into aquifers. A World Bank report revealed that poor water quality threatens growth, harms public health and imperils food security.[3] For example, severely degraded water quality eliminates two third of potential economic growth, and groundwater salinity diminishes agricultural productivity. Every year, salinization of freshwater sources eliminates enough drinking water to sustain 170 million people per day.

Most water quality and water quantity data are captured at a highly localized scale, making it difficult to understand the status of groundwater at a national scale. Lack of national-scale characterization limits implementers' ability to develop large-scale programs. A deficit of groundwater contamination knowledge ultimately affects public health outcomes. For example, this is observed in the case of fluoride contamination in the Rift Valley in East Africa. Limited data are available on the scale of fluoride-contaminated groundwater, which potentially impacts the health of many residents.[4]

The absence of succinct hydrological knowledge affects overall water security. When hydrogeology is not well understood, wells can become less productive or dry up. Implementers may not be able to track or predict groundwater production, thus jeopardizing marginalized communities.

DECISION STATUS QUO

Decision-makers typically rely on field sampling to evaluate water quality or measure water quantity. If they are unable to go into the field, they have to refer to previous reports to identify critical areas. To anticipate climate change, decision-makers have to rely on scientific predictions and reports, publications, or recommendations. Outdated data and generalized reports or predictions limit how well a response can support safe water management, including at small spatial scales.

Uncertainties of climate predictions, in combination with the fact that climate models are highly complex, prevent governments and CSOs from anticipating future groundwater crises and potential limit their ability to implement sustainable programs.

DEMAND

Conversations with various stakeholders revealed high demand for increased availability of high-quality groundwater data and predictive models. A conversation with the Director and a Researcher at the International Groundwater Assessment Center (IGRAC) indicated a "need for data on groundwater quality at a regional level."[5] IGRAC focuses on transboundary aquifer assessments and groundwater monitoring. The Director also noted, "We need more awareness of the importance of groundwater quality."

A growing part of the United Nations Children's Fund (UNICEF) funding is tied to climate change adaptation and climate change mitigation. One of their areas of interest is climate change resilience and vulnerability of water supplies. A UNICEF representative also highlighted the importance of making data actionable.[6] UNICEF is seeking data on groundwater quality, depletion, water scarcity, and impacts of climate change.

The German Federal Institute for Geosciences and Natural Resources (BGR), provides independent advice on natural resources issues. They have a specific department focused on groundwater, including exploration and protection and quantitative and qualitative groundwater assessment. BGR expressed disappointment in the lack of groundwater data, “Compared to surface water, there is not much data on groundwater. Data on [groundwater] quality is even more scarce than on quantity.”[7]

The World Bank in Ethiopia showed interest in how to collect data in a more effective way, but they highlighted the absence of staff to analyze data at a government level.[8] The interviewee pointed out a data deficit in resource availability and resource quality in Ethiopia (illustrated by wells that dry up in Ethiopia and concerns of fluoride contamination). The World Bank more broadly indicated an interest in technology to improve resource management, “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.”[9]

OTHER DATA APPLICATIONS

Application 1: Predicting water conflict

Decision: Outputs from applications can help governments and CSOs anticipate future conflicts, based on quantity forecasts. The better we can predict these conflicts, the better the decision makers can prevent these conflicts by responding to the local issues and by bringing the help that is needed.

Demand: The team from the Water, Peace and Security (WPS) partnership developed a machine-learning model for forecasting conflict up to a year in advance.[10] They tested more than 80 indicators; water variables such as precipitation anomalies, flood risk, and seasonal and interannual variability were among the 20 most relevant indicators. Predictions that include climate change and forecasted water access could improve the ability to predict conflicts.

Application 2: Predicting population displacement

Decision: Outputs from applications can help anticipate population movement and future migrations related to climate change and deterioration of water resources.

Demand: Expected population resettlement due to climate change would be exacerbated by the reduction of available water resources.[11] It is critical to better understand the balance between water availability and need, to limit population displacement due to water scarcity.[12]

Application 3: Optimizing irrigation infrastructure

Decision: In conjunction with groundwater quality forecasts, agricultural stakeholders could use applications to optimize the location of irrigation infrastructure and the amount of water that has to be withdrawn, so as to maximize crop yields and minimize environmental degradation.

Demand: The World Bank reported two major implications of having a degraded water quality.[3] The first one is related to the loss of crop productivity due to irrigation with saline water. The second major implication is using untreated wastewater for irrigation, which threatens the health of 885 million urban residents in low- and middle-income countries.

EXISTING AND UPCOMING INNOVATIONS

Innovation 1: Models from researchers at the University of California, Davis and the United Geological Survey (USGS)

Relevance: Existing model developed to predict nitrate and arsenic in groundwater in the United States

In an effort to address water demand and agricultural needs in the Central Valley of California, researchers developed, "a hybrid, non-linear, machine learning model within a statistical learning framework to predict nitrate contamination of groundwater to depths of approximately 500 m below ground surface."[13] The model used 145 predictor variables related to water, nutrient balances, land use, soil, climate, etc. They used a boosted regression tree method to identify the most relevant variables. This model is a direct example of the use case.

Researchers within the National Water-Quality Assessment program of the USGS has developed a model to analyze nitrate and arsenic in the southwestern United States. They used a "random forest classifier algorithm to predict concentrations…across a model grid."[14] The model used natural and anthropogenic variables to ultimately predict aquifer vulnerability to contamination.

These types of models are numerous and useful, but are often underused, hard to interpret, or limited to locations in developed countries.[15,16]

Innovation 2: Groundwater modeling with machine learning

Relevance: Existing model developed to predict groundwater levels in Ljubljana polje aquifer (Slovenia)

Researchers in Slovenia and Greece conducted an analysis of machine learning methods to forecast groundwater quantity in the Ljubljana polje aquifer.[17] They tested numerous models to identify the best-fitting method. In the end, they were able to predict groundwater levels successfully and compare against field-collected data. This work is a great example of future opportunities in forecasting groundwater quantity.

Innovation 3: Digital platform that aggregates and provides quality control of groundwater data from any source.

Relevance: An existing model for coalescing surface water data will be adopted for groundwater.

In California, the State Water Resources Control Board aggregates surface water quality data provided by any stakeholder via the California Environmental Data Exchange Network (ceden.org) to facilitate regulatory decision-making and public access.[18] IGRAC has implemented a participatory groundwater data network called the Global Groundwater Monitoring Network with support from the United Nations Educational, Scientific and Cultural Organization (UNESCO).[19] In partnership with Aquaya and others, IGRAC aims to produce a new global data aggregation portal (groundwater-quality.org) that would facilitate large-scale assessment of groundwater.[20]

Innovation 4: Digital platform that offers data treatment services to help organizations develop their activities

Relevance: Example technology could be redeveloped or enhanced to meet the use case needs.

The Global Water Intelligence platform is an online tool that displays data on the water market by industries or by countries.[21] They offer partnerships to connect water sector actors. SERVIR provides visualization tools to better manage climate risk, natural disasters, food security, water resources, and land use change.[22] Earth Genome provides solutions to address water scarcity through different tools available on their platform.[23] The India Water Tool is a platform for companies, CSOs, and governments to identify the water risks across a country.[24] Users can access data on groundwater level and groundwater quality.

PARTICIPANTS

National governments (e.g., ministries of water): National government representatives would likely be at a management level at a water ministry. They understand reports and briefs (but not scientific articles) and can interpret maps. They don't have skills to manage data but can understand statistics. Their time is limited, so they need a straightforward software application that has simple processes (i.e., downloadable reports and maps). National government representatives would use applications (and downloaded reports and maps) in the following ways:

- Develop a legal framework to protect water resources (quantity and quality).
- Identify critical areas where the water resources need to be more protected (protect the groundwater recharge areas).
- Create action plans addressing anticipated effects from climate change.
- Develop plans to address water issues in poorer areas with more vulnerable people.

Local governments (e.g., regional offices for a water ministry): Local government representatives would likely be at a management level within the regional office. They have less experience in understanding data than the national government representatives, but they can download reports and interpret maps. They have more time and could use a simple software application to help them understand the state of resources locally. Local government representatives would use applications in the following ways:

- Identify poor water quality areas within their region and work with community groups to identify solutions.
- Identify communities with poor access to safe water within their region and plan appropriate interventions to address access issues.

Civil society organizations (CSOs): Project managers within CSOs likely have more experience working with data and interpreting results. Although smaller CSOs might not have enough resources to manage data independently, they can understand reports, maps, and statistics and can use applications to understand the situation on the ground. Larger CSOs may have dedicated staff to manage data, who are able to treat raw data and use applications to extract maps or perform modeling. CSO project managers would use applications in the following ways:

- Identify critical areas in which to prioritize their programs.
- Understand water quality or water access needs at a regional or national level and tailor their program design accordingly.
- Anticipate areas with possible future crises (flood/drought) and target programming in these areas.

Donors: Program managers at donor organizations typically have minimal skills in data management and interpretation (outside of financial figures). They often have limited time to spend understanding data when determining funding priorities. They can understand reports and basic statistics and interpret maps. Funders would use applications in the following ways:

- Interpret the geographical areas that need more attention as related to water quality or water access, to direct funds to these locations.
- Anticipate future funding needs and develop strategies that correspond to these needs.

Agricultural stakeholders: Agricultural stakeholders using applications would be at a management level within their organization. They would be very familiar with agricultural data, understanding statistics, crop yields, etc. They would have some time to interpret applications given their potential to increase profits. Agricultural stakeholders would use applications in the following ways:

- Develop and download maps that indicate optimal locations for irrigation infrastructure.
- Model the ideal amount of water that has to be withdrawn, so as to maximize crop yields and minimize environmental degradation.

Developers: Developers are individuals (or teams) that develop various tools. Developers have advanced computer programming and design skills. They are familiar with modeling, maps, and how to iterate development to suit participant needs. Their time would be devoted to application development during the initial phase, followed by maintenance and upgrades.

OUTPUTS

The envisioned output for this use case would have the following characteristics:

- Groundwater forecast data will be accessed through a web platform and visualization dashboard. Users will be able to search the data they are interested in using keywords, along with other filtering parameters.
- Data will be downloadable as raw data (Microsoft Excel files) and as maps (e.g., predictive, risk).
- Reports will be developed in conjunction with groundwater forecasts, to describe areas of highest vulnerability.

WORKFLOW

Step 1: Gather data

Collect and list all data (up to a global scale), including water quality and quantity, climatic data (precipitation, temperature), environmental parameters (e.g., land cover, elevation, rivers, lakes), and socio-economic parameters (e.g., population density, poverty, access to water, access to sanitation).

Step 2: Website development

Develop a user-friendly website that allows users to find data by type or by geography. Data will have to be cleaned and converted to a format that is easily accessible. Data will be uploaded and well-organized.

Step 3: Data visualization

Developers will need to create an interactive platform for users to visualize data. This process will need to be iterative with users' feedback.

Step 4: Predictive models

Users will build groundwater quality and quantity models using available data. They will be able to produce processed data and maps as outputs from the models. Standard models can include preset variables with relevance to climate change, economic development, and agriculture.

DATA SOURCES

Existing data:

- Multiple Indicator Cluster Surveys (MICS) are publicly available household survey data produced by UNICEF using internationally comparable, statistically rigorous methods.[25] Unfortunately, the geographic coordinates are not publicly available. As of 2018, UNICEF was working on including actual field water quality testing as part of the dataset.[26] More than 30 countries have been updated with integrated water quality data.[27]
- Various implementing organizations such as multilateral organizations, CSOs, and utilities collect water quality and quantity data throughout their program implementation and research. However, this information is often only available as summary data at a national level, or is not publicly available.
- Research institutions collect high-quality water data (both quality and quantity), but often they are not shared publicly.
- Climate models are becoming more widespread, but often the models are propriety, highly complex, or not easily accessible.

Data gaps:

- There is limited expansive data on water levels and water depletion, which is necessary to understand areas of future risk.
- Groundwater quality data are scarce and difficult to access, and typically only available at small scales. There are many methods of data collection and many parameters to consider. Institutions that are collecting water quality data are often focused on one or two specific parameters and might not have the capacity or funds to collect more parameters. In addition, it can be challenging to monitor sampling and analysis execution quality.

BARRIERS

Accessing existing data: Data collection will be challenging since the data are sparse and are collected by many different organizations. A thorough review of all organizations that collect data (whether required or voluntary) is needed. Facilitators need to work with each organization regarding privacy of their data, to ensure their data are available to use in application development.

Additional data collection: There is a real need to collect more data on both groundwater quality and quantity. Going into the field is expensive and time consuming, especially in remote areas.

Spatial variability: There is a great deal of spatial variability in existing data. Integrating data across geographic and political boundaries will be a challenge.

2. WATER: REDUCING NON-REVENUE WATER

OBJECTIVE

To provide water suppliers (“utilities”) with real-time data on flow and pressure within the network to reduce non-revenue water.

Decisions:

- Utilities could use real-time data on flow and pressure within water distribution networks to identify leakages, pipe bursts, illegal connections, and meter disruptions to decide where to direct remedial actions, such as pipe or meter replacement, pressure management, or connection removal.
- Utilities can also use these data to anticipate peaks in consumption and adapt water production accordingly.

DESCRIPTION

Non-revenue water (NRW) is a chronic challenge facing water utilities globally. One of the major issues affecting water utilities in the developing economies is the considerable difference between the amount of water put into the distribution system and the amount of water actually billed to consumers. In some low- and middle-income countries, about 35% of water is lost daily.[28] This water loss is defined as non-revenue water (NRW), or water that is pumped and then unaccounted for.[29] NRW has two main components: i) physical losses (e.g., leakage from pipes and storage tank overflows), and ii) commercial losses (e.g., under-registration of customer meters, data handling errors, and illegal connections or theft). Physical losses comprise leakage from all parts of the system and overflows at the utility's storage tanks. They are caused by poor operations and maintenance and/or the lack of active leakage control. Collection of real-time flow and pressure data helps to inform pressure adjustment and pipe replacement.

A third source of NRW is unbilled authorized consumption, which includes water used by the utility for operational purposes, water used for firefighting, and water provided at no cost to certain groups. However, unbilled authorized consumption does not reflect operational inefficiency, but rather a public policy decision to allocate water without monetary compensation.[30]

In low- and middle-income countries, utilities collectively lose roughly 45 million cubic meters of water per day, which amounts to a value of more than 3 billion USD per year.[28] Water utilities spend large sums of money on treating and pumping water intended for customers that becomes lost. Poor understanding of the magnitude, sources, and cost of NRW is one of the main reasons for insufficient NRW reduction efforts around the world.

DECISION STATUS QUO

Most water utilities have inadequate monitoring systems to assess water loss. They typically identify repairs through staff observation or reports by concerned members of the public. In addition, many utilities lack real-time data on flow and pressure. Network monitoring is limited as it is typically done through in-person readings of analogue meters. When available, flow and pressure measurements usually have sparse spatial and/or temporal resolution. Without high-frequency measurements in several strategic locations throughout the network, utilities cannot: i) perform a water balance calculation and estimate physical losses in different sub-sections of the network (often called "district metered areas"), or ii) detect changes in "minimum night flow," which is a standard technique for identifying physical losses. Utilities may collect consumption behavior data, but not understand how to improve water regulations to account for fluctuations in water demand. **Without granular data on NRW, utilities may be unable to improve operational efficiency, meet customer demands, or achieve sustainability.**

DEMAND

All water utilities expect some physical and commercial losses. The first step for any utility aiming to minimize water losses is to prepare an NRW audit to establish baseline levels of water losses. “First, do a water audit to identify the proportion of commercial and physical losses. Then prepare a plan, [including a] financial plan.”[31] However, this step is missing for many utilities. “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.”[31]

Once an audit is performed, the goal is to reduce losses due to non-revenue water. Demand for reducing NRW is high. In 2016, the World Bank and International Water Association (IWA) established a global partnership to help countries, especially the poorest, to address NRW. The program aimed to promote the use of performance-based contracts to address NRW. It also engaged with regional development banks to facilitate financing, streamline the preparation of performance-based contract (PBC) transactions, and increase the number of stakeholders active in the market.[32,33] In 2019, the Public-Private Infrastructure Advisory Facility (PPIAF) allocated 300,000 USD to support the Water Services Regulatory Board (WASREB), the independent water sector regulator in Kenya, in promoting PBCs in NRW reduction. In another example, the Government of Vietnam received 330,000 USD to reduce NRW in Vietnamese water utilities to improve energy efficiency across the sector.[34]

Kenya provides one example of national-level programs to address NRW, particularly through performance incentives. In 2014, WASREB published standards for NRW management. The standards provide a basis for addressing challenges and suggest measures to reduce losses, such as monitoring systems at production, distribution, and consumer levels. In addition, WASREB monitors the performance of the approximately 90 regulated urban Water Service Providers (WSPs) against nine key performance indicators (KPIs), which specifically include reduction of NRW. The goal of WASREB’s National Water Services Strategy is to reduce NRW to less than 30%, while the Kenya Vision 2030 goal is to reduce NRW to less than 25% (“acceptable”); the “good” sector benchmark for NRW is 20%.[35] WASREB introduced a tariff system that allows utilities to secure a budget for NRW reduction measures through their revenues. Tariffs are linked to the achievement of KPIs and service levels, such as water quality, water supply, and reductions in NRW.[35,36] Nine utilities in

Kenya have already started NRW reduction programs.[35]

OTHER DATA APPLICATIONS

Application 1: Understand customer consumption behavior

Decision: Outputs from applications could provide water suppliers with data on historical customer consumption patterns to forecast water demand and guide an improvement strategy.

Demand:

- “Sometimes you start engaging with a utility, and the most fundamental basic thing that you want to know is […] ‘Do you know who your customers are?’ And the utilities have no idea. A lot of these utilities are lacking a fundamental understanding of their own systems, so it’s hard to build an improvement strategy.”[8]
- “The data […] estimates range from meters not read because nobody opened the gates of their homes to the meter reader, so the meter readers were not able to locate the homes. These 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.”[37]

Application 2: Identify priority targets for NRW reduction programs

Decision: Robust data on water supply performance combined with advanced analytical methods can help to identify where NRW interventions would be most cost-effective.

Demand:

- "[Non-revenue water] is easier to fix in smaller utilities."[31]
- Generally, the Water Sector Trust Fund in Kenya is interested in water service level data to target investments. "It is important to determine the levels of service, so that the Water Sector Trust Fund knows who to target. This information could help understand what kind of investments [we] should make in specific areas (e.g., not only where to support, but also how/what)."[38]

EXISTING AND UPCOMING INNOVATIONS

Innovation 1: Remote monitoring technology

Relevance: Remote monitoring allows users to collect data from a distributed network of field sensors without needing to visit each location. Remote metering technology helps to track water at the household meter level. These upcoming technologies could enhance the ability to achieve the use case objective.

Examples: Metasphere Point Colour RTUs provide data on water pressure and flow to help WSPs identify and detect leakages and bursts within their networks.[39] The Point Colour RTUs usually power up the sensor and take regular readings every 15 minutes. RTU data are transferred back to a central location where clients can view a management dashboard (such as EcoStruxure™ Geo SCADA Expert,[40] Mosaic,[41] AVEVA OSIsoft,[42] or Palette,[43] Metasphere's own telemetry platform), typically once a day.

CityTaps has developed smart water meters (CTMeter) that send water usage data in near real time to the cloud-based software, CTCloud.[44] In addition to tracking payments, the CTCloud and meter systems provide utilities with information on functionality and data analytics, to help identify issues. CityTaps uses a prepaid meter technology; however, there is potential to develop software to identify leaks (e.g., above-average flows for a sustained period of time), meter tampering, and water theft. Water consumption data from CityTaps meters can also help to forecast demand peaks.

Innovation 2: Cloud-based Supervisory Control and Data Acquisition (SCADA)

Relevance: SCADA is an existing technology that can contribute to achieving the use case objective (provision of real-time data on network flow and pressure).

Cloud Supervisory Control and Data Acquisition (SCADA) systems, a recent advance from traditional network-based SCADA, combines remote monitoring with the control of field-based equipment (e.g., pumps, valves, and treatment equipment) from a central location. XiO is an example of a company that develops cloud-based SCADA systems exclusively for the water, wastewater, and irrigation applications.[45] In the case of water distribution networks, remote telemetry units (RTUs, or electronic monitoring devices) can monitor and report flow and pressure remotely through wireless communication technology. Their installation allows real-time monitoring of District Monitoring Areas (DMAs), which assists workers in managing and repairing pipe bursts and leaks as soon as possible. Continuous pressure monitoring can also identify high pressure areas for pressure reduction and highlight issues related to pumping.[45]

Innovation 3: Artificial Intelligence (AI) technology

Relevance: Artificial Intelligence (AI) algorithms trained to identify pipe bursts using flow and pressure data. AI algorithms are being developed and refined to achieve the use case objective.

AI deals with simulation of intelligent behavior in computers. AI can be applied to pipe monitoring and leak detection. For example, numerical algorithms developed for water leak detection aim to detect certain spatial and temporal patterns and anomalies in flow and pressure values at different points in the water distribution network. These data can then be used to extract information on physical and commercial losses. AI algorithms can also classify water losses as illegal connections, water leaks, pipe bursts, malfunctioning sensors, abnormal water consumption patterns, etc.[46]

Innovation 4: Remote sensing (satellite) technology

Relevance: Remote sensing technology such as Hydro-Scan Technology from Utilis detects soil moisture belowground, and thus can potentially identify pipe leaks/bursts.[47] Technology such as Leakmited recently achieved proof of concept by detecting leaking water pipes using satellite measurements of the Earth's electro-magnetic back scattering. These existing technologies can contribute to achieving the use case objective.

Hydro-Scan technology uses radar data, taken from a satellite, to detect soil moisture up to 10 feet belowground. The technology was initially marketed as a construction aid to determine where pavement might fail due to soil moisture; however, it can also be used for leak detection. Using an algorithm, Utilis can analyze the satellite data to determine the signature of drinking water and therefore detect leaks. Due to the nature of the satellite data, the analysis is unaffected by day/night light and weather issues. Although the existing system does not allow for real-time data, surveys are conducted every two weeks. Combined with real-time flow and pressure data, this technology could help identify leaks much more efficiently. The technology is completely remote and accompanied by user-friendly geographic information system (GIS) reports.

Utilis completed a case study in Bangkok, Thailand, in February of 2020. Utilis worked with the Metropolitan Water Authority of Bangkok over two months and detected more than 2,000 leaks, which is the equivalent of 390 L/s, or enough water to supply 100,000 people. They corroborated leak detection with on–the-ground verification and found that 90% of the satellite-detected leaks were actual leaks.[48]

Leakmited is a company that provides direct services to utilities.[49] Their goal is to reduce NRW by finding leaks and predicting which pipes are vulnerable to leaks. A client provides Leakmited with the GIS files of their existing system and infrastructure and historical records of pipe failures and leaks. Leakmited processes the data using algorithms, in conjunction with satellite imagery, and provides the client with possible leak locations. Clients must validate leaks on the ground.

Innovation 5: Leak detection hardware

Relevance: Newer methods for leak detection take advantage of technological capabilities. For example, Flexim is a non-invasive ultrasonic flowmeter for underground pipeline monitoring and leak detection.[50] The Sahara Leak Detection Platform can be used to accurately locate and pinpoint leaks on primary and secondary main distribution pipes.[51] Augmented reality visualization platform, vGIS, is an upcoming virtual reality technology. Such existing and developing technologies can contribute to achieving the use case objective.

Flexim's non-invasive ultrasonic FLUXUS flowmeter provides flow measurements to identify water losses from underground drinking water lines. One main advantage is that the installation of the measuring system does not disrupt operations (i.e., does not require opening pipes); the transducers are attached to the outside of the pipe wall. The FLUXUS ultrasonic flowmeter detects small volume flow rates with high precision. With flow measurement data points, leaks can be pinpointed. Data are transmitted to the grid control system.

WSPs usually establish a leak detection team responsible for locating non-visible leaks on primary and secondary mains (generally pipes >300mm in diameter). A WSP may want to use the Sahara Leak Detection Platform to locate and pinpoint leaks within a sub-area of the network suspected to have a leak following interpretation of flow and pressure data. The Sahara Leak Detection Platform is a tethered tool with live video that can accurately identify leaks and air pockets in water and wastewater pipelines. It is inserted into the pipeline and propelled by flow velocity. If the tool encounters an acoustic event, such as a leak, the operator can stop it at the exact point of the leak. At the same time, the team can locate the sensor and mark the exact leak location (within a 0.5-meter range).

vGIS was developed to help municipalities and utility companies visualize underground networks to prevent excavation-related accidents.[52] Although municipalities and utilities often maintain good records of existing infrastructure, personnel in the field cannot necessarily access or easily interpret those records. vGIS uses augmented reality visualization to produce holographic images of underground networks, including pipes, valves, cables, and other utility objects. vGIS can be updated in real time. Future applications could combine the visualization with remote sensing technologies and pressure and flow data to identify pipe leaks and help utilities quickly identify pipe locations that need repairs.

PARTICIPANTS

Service providers – water utilities: Water utilities are the primary users of the technology. Within utilities, there are two types of primary users:

- Supervisor/manager: Supervisors or managers ("managers") are the primary decision makers as related to the objective. Managers have likely been with the utility for several years and are familiar with the utilities' infrastructure. They likely have a background in engineering and should be very comfortable with data, data analytics, and interpreting maps and planning drawings (architectural/engineering drawings). Expertise may depend on hiring practices. In some cases, managers might rely on internal data analysts or GIS specialists to help interpret data. Their primary job is to ensure smooth operation of the utility and they will have time to dedicate to understanding real-time data. Managers would use applications in the following ways:
    - Using real-time data, designate field personnel to address any existing challenges within the utility (e.g., burst pipes, illegal connections, disconnected meters).
    - Prioritize how to address various challenges, know which field personnel are available, and understand the available and required resources.
- Field personnel (or leak detection teams): Field personnel are the individuals or team reacting to the situation on the ground, as delegated by managers. Field personnel are familiar with the utility infrastructure, have construction experience, and are somewhat comfortable with planning drawings and data. Their time is dedicated to fixing repairs, disconnecting illegal connections, monitoring meters, and communicating with customers or would-be customers. When they are in the field, it can be difficult to have the correct drawing on hand and interpret it accurately. Field personnel would use applications in the following ways:
    - Using maps and drawings provided by managers, they repair any burst pipes, illegal connections, or disconnected meters.

Developers: Technology developers are individuals (or teams) that develop a system/application to meet the objective. Developers have advanced computer programming and design skills. They are familiar with systems networking, data platforms, virtual reality, GIS, etc. They are comfortable with the technology development process and know how to iterate technology development to suit participant needs. Their time would be devoted to the development of the technology during the building phase, followed bymaintenance and upgrades.

OUTPUTS

- Real-time data on flow and pressure can be developed and maintained on internal utility applications. Utility managers can be trained to interpret data and record actions taken to maintain a database.
- Mapping software can be used to easily track the existing network, areas of concern, and customer connections.
- The software will be used to produce standard or custom reports for upper management or regulatory agencies. Reports can also include consumption patterns.

WORKFLOW

Step 1: Establish a water balance

The first step in reducing NRW is to map the pipeline network and establish baseline levels of water losses through an audit (i.e., using available information supplemented with targeted field measurements and data). Audits are further described in steps 2 and 3.

Step 2: Conduct field audits

Field audits consist of field surveys in selected parts of the service area to obtain information on water delivery status. This includes assessing the status of the delivery network, surface leaks, and illegal connections. Field audits are typically undertaken by a team of technicians.

Step 3: Conduct commercial audits

Commercial audits consist of a detailed survey of the water utility customers in a specified zone or area. The purpose of the survey is to obtain relevant customer data, water usage, and the status of water services, including water meter function. Commercial audits can be used to provide advice and guidance to customers on water security and the good water use practices, limiting waste.[28]

Step 4: Form DMAs and install sensors

The creation of DMAs is key to managing NRW. A DMA comprises a discrete area within the distribution network, typically serving 500–5,000 connections, with meters on all inflows.[28] The size of the DMA is determined on a case-by-case basis and depends on a number of factors (e.g., hydraulic, topographic, economic). DMAs allow a water balance to be derived at a granular level so that levels of physical and commercial losses can be assessed and targeted effectively.

Once formed, staff can compute the volume of leakage in each DMA (i.e., hydraulically discrete zone) with water meters monitoring flows in and customer meters monitoring flows out. This allows leak detection specialists to better target their efforts. A simple water balance can be used to understand leakage and other losses within the DMA. With DMAs across a whole water supply network, utilities or operators can identify high loss areas and then target pipe replacement/repair, leakage, and pressure management projects to reduce NRW.

Step 5: Install sensor technologies for pressure, flow, and leak management throughout the piped network

The next step is to install technology (e.g., XiO SCADA system, FLUXUS flowmeter) for water network monitoring, such as pressure, flow, and leak management. For SCADA systems, real-time data on parameters of interest from RTUs is sent to the cloud. For RTUs, data (i.e., sensor readings) are transferred to control systems and trigger alarms if a sensor measurement is outside an established threshold.

Step 6: Monitor monthly and yearly flow and pressure data

After several months or years of flow and pressure data are collected, it can be compared with actual leak events to train an AI algorithm to recognize leaks. In parallel, after several months or years of water consumption data are collected, AI can be applied to forecast consumption. AI algorithms learn and improve as more data become available to forecast water demand at a node (point of flow withdrawal) or group of nodes. Demand can be produced in real time for the next 24 hours or longer term (years).

Step 7: Establish key performance indicators

Standardized performance indicators should be calculated according to a clearly defined methodology and using standard definitions, to enable performance comparison.

Step 8: Sustain NRW reductions

System maintenance, including network maps, sensors, and algorithms, is critical. Without maintenance, NRW losses can increase over time.

DATA SOURCES

Existing data:

- Many utilities manually collect data on pressure/flow. However, since the data are manually collected, they are not available in real time. In addition, utilities generally conduct minimal analysis on this data.
- Utilities may maintain customer level data, including contact information, address/location, and billing information.
- Most utilities maintain a database of existing pipe networks, whether drawn by hand or digitized. However, these databases might not be well-maintained, complete, or up to date.
- Utilities may record known leakages.

Data gaps:

- Real-time flow and pressure data.

BARRIERS

Institutional barriers: Institutional barriers, such as weak regulatory enforcement, operational inefficiency, and lack of understanding of the issue will pose challenges.

Cost-effectiveness: Finding the most cost-effective solution for WSPs relies on institutional buy-in from senior leadership. Senior leaders often do not recognize NRW as a significant issue.

Attracting finance: It can be difficult for utilities to obtain funding to reduce NRW, since there is a perception of financial risk with minimal impact.

3. SANITATION: COORDINATING FECAL SLUDGE EMPTYING

OBJECTIVE

To provide governments (local or national), the private sector, and customers a application that optimizes urban fecal sludge emptying activities. It should: i) track available geo-referenced sanitation pit-emptying jobs, ii) suggest routes and pit locations that should be targeted (due for emptying soon), and iii) maps historical emptying jobs (frequency of desludging events).

DECISIONS

- Governments could use an application to accurately track where service providers are releasing fecal wastes, and thus decide how to improve disposal regulations. Data from this application could also provide governments with better information to decide where to focus future investments in fecal sludge management (e.g., where to extend existing sewer networks, how to reach areas that are difficult to access, where to locate new treatment facilities).
- Service providers could use an application to improve knowledge about when to visit which pit latrines/ septic tanks, where to dispose of fecal sludge, and to whom to market their services. Improved decisions should reduce overall costs and increase company efficiency.
- Customers could use an application when to easily request emptying services for their pit latrine/septic tank, identify the appropriate service provider, compare service provider prices, and ensure their waste is legally and safely managed.

DESCRIPTION

As governments strive to achieve SDG 6.2, data on fecal sludge emptying practices and services could play a role in improving public health and environmental outcomes. The WHO/UNICEF JMP's definition of safely managed sanitation specifies that fecal waste is either transported through a sewerage system for off-site treatment or temporarily stored in a safe on-site containment structure prior to transport and off-site treatment.[53] In urban areas with high population density, such as low-income areas, treating and disposing of fecal sludge in situ is challenging. Further, extending sewer networks to these areas is often costly due to infrastructure challenges. Therefore, one of the only practical solutions is to empty fecal sludge mechanically (e.g., using vacuum trucks) or manually.

In many urban areas in low- and middle-income countries, emptying services are performed on an ad-hoc basis. Waste emptiers may belong to an association of providers, but there are limited methods for customers to get in touch with service providers and little initiative on the part of service providers to reach out to previous customers for continued services.[54,55] Although professional associations may be formalized, they often lack data on how their business is working. For instance, "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."[56]

Governments are tasked with regulating fecal sludge emptying service providers.[54] However, regulators are unable to track all fecal sludge dumping events. One interviewee noted, "If trucks and dumping sites were geo-localized so that they could be tracked on a dashboard, it would help fight wild dumping."[57]

Examples from Kenya and Bangladesh highlight the fecal sludge emptying challenges faced in urban areas. In Kisumu, Kenya for example, approximately 60% of the population uses unimproved sanitation.[58] Overall, in Kisumu city, only approximately 20% of residents are connected to sewer, with 5% connected to septic tanks.[59] The majority of Kisumu low-income residents 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 or rivers, contaminating the environment and posing a serious public health risk.[60]

DECISION STATUS QUO

Residents of low-income, urban settlements in low- and middle-income countries do not generally have access to formal, regulated fecal sludge management (FSM) services, which primarily target middle- and high-income households that can afford market prices.[61] FSM in poor neighborhoods commonly includes unsafe practices, such as employment of informal manual emptiers who remove fecal sludge by hand and dispose of it in the surrounding environment (e.g., burying it onsite or disposing of it in nearby waterways). Increasing the use of safe and regulated emptying services is therefore critical for improving sanitation safety in urban areas. Governments, service providers, and customers can continue to make decisions using the usual approaches. **Without the ability to improve decision-making, however, there will be limited improvement in the safe management of fecal sludge in dense, urban environments.**

DEMAND

Challenges inherent in existing pit-emptying systems are illustrated in research conducted in Port Harcourt, Nigeria. Potential customers 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."[62] Meanwhile, vacuum truck operators struggled with enough work to fill their day: "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."[63]

Similarly, a manager at a Kenyan water and sanitation facility indicated, "This [application] will help the utility to know how many toilets and the amount of sludge directed into their sewerage network, the number of onsite sanitation facilities that can be emptied by exhausters and those that can be emptied through manual emptying."[57]

A application could also address the constraint of not understanding existing sanitation infrastructure. "Onsite sanitation is very poorly monitored in most countries (hardware and services); e.g., type of sanitation coverage, who empties tanks, how often tanks are emptied, where fecal sludge goes. There is miscategorization of containment systems."[64]

In another example, Sanergy (Nairobi, Kenya) established a franchise system for community-level container-based sanitation in a low-income area of Nairobi, Kenya. Fecal waste from toilets is removed by handcarts and trucks following an emptying schedule established when the toilet is installed. The emptying schedule, however, does not account for seasonal patterns (e.g., heavy rains). Sanergy could benefit from an application that notifies emptiers when a given pit is full.

OTHER DATA APPLICATIONS

Extended uses of the proposed sludge-emptying application might include the following:

Application 1: Develop a database of all existing service providers

- **Decision:** Governments (local or national) could use a database of all existing service providers in conjunction with notifications of illegal discharge, to determine the best methods to optimize illegal discharge, and formalize the sector through relevant licensing and regulation.
- **Demand**: Onsite sanitation is very poorly monitored in most countries, for example, there is limited information regarding the type of sanitation coverage, who empties tanks, how often tanks are emptied, where fecal sludge goes. There is also mis-categorization of containment systems. For example, septic tanks are often not actually true septic tanks; without a leach pit, the liquid effluent is a contamination risk."[64] In Kenya, the independent regulator, WASREB, aims to "get data on sanitation service providers (e.g., VTOs and manual emptiers) [as] data is supposed to support in the decision-making process."[65]

Application 2: Develop a database of existing sanitation infrastructure (locations, quality of facilities, type of facilities)

- **Decision:** Governments (local or national) as well as implementing organizations (e.g., CSOs, private companies) could use a database of existing sanitation infrastructure at a granular level to focus sanitation interventions on areas of greatest need and prioritize opportunities for fecal sludge reuse.
- **Demand:** "It would be extremely useful to have data on the mapping of toilet locations (GPS-marked) and type of toilet in all informal settlements. This would allow for a summary of sanitation infrastructure for each informal settlement. This could also inform Sanergy's expansion into other informal settlements and [their] engagement strategy and sales conversations. If Sanergy understands the dominant toilet type with in an informal settlement, [they] can frame their marketing dialogue to engage the most households as possible."[66]

Application 3: Predict characteristics of fecal sludge, using the database of existing sanitation infrastructure with agricultural activities and additional variables

- Decision: Governments (local or national) as well as utilities and researchers, could use the database of existing sanitation infrastructure in conjunction with agricultural activities at a granular level, to predict characteristics of fecal sludge and thus design appropriate treatment facilities.
- Demand: "Designing and operating new treatment plants requires an extensive data collection cycle beforehand and [sludge] characteristics vary depending on the season or samples from different households [or types of infrastructure]."[67] This extended application could reduce how much additional data is needed to predict fecal sludge characteristics and thus design appropriate treatment facilities.

EXISTING AND UPCOMING INNOVATIONS

Innovation 1: Pula mobile application, Sub-Saharan Africa

Relevance: An existing mobile application for service providers that can be applied to achieve the use case objective.

Pula is a mobile application developed by Water & Sanitation for the Urban Poor (WSUP) over a 3-year period to address service providers' lack of access to data on their customer base, operating standards, and levels of service.[68] WSUP investigated needs of service providers in four countries and identified two key features needed for the application to be successful: i) an address book for service providers to save customer information and track when their facilities need additional servicing, and ii) a tracking feature to share truck locations with business owners. Initial piloting in Lusaka, Zambia and Maputo, Mozambique revealed that the application was too complex for the intended users.[69]

WSUP identified lessons learned from the trial of Pula that can be used in the development of future applications:

1. “Focus on one core feature and ensure this is fit for purpose;
2. Establish a relationship with one target customer, allowing the product to be tested over longer periods; and
3. Focus on developing a product tailored to a single market, which can then be adapted for new markets as required.”

Innovation 2: Mobile applications, Uganda (Kampala Capital City Authority) and Bangladesh (Practical Action)

Relevance: An existing mobile application for pit-emptying service delivery that can be applied to achieve the use case objective.

With support from the Global System for Mobile Communications (GSMA) Mobile for Development (M4D) Utilities Innovation Fund, two separate applications have been developed to connect pit-emptying entrepreneurs with customers and track service delivery across the sanitation value chain.[70] The applications offer customers in low-income areas access to appropriate emptying service, and allows the municipality to map sanitation facilities and track and coordinate regular pit-emptying activities. The municipalities have access to a database of sanitation facilities with characteristics, frequency of emptying, and distances between facilities and treatment plants. In Kampala, the municipality hopes to use the data from this application to guide investments, allocate resources, regulate service delivery, and enforce standards. In Bangladesh, the application also allows for mobile payments, thus speeding up the transfer of funds.

Innovation 3: GIS service optimization tool for fecal sludge collection, Thailand

Relevance: An existing tool for fecal sludge management that can be applied to achieve the use case objective.

In Nonthaburi Municipality, Thailand, researchers logged GPS data from fecal sludge collection trucks over six months to produce two algorithms.[71] The first algorithm optimized the grouping of emptying activities to minimize travel time. All emptying activities were assumed to have equal priority. The included trucks required returning to a treatment plant after every three emptying activities. The second algorithm prioritized emptying locations that could not be emptied on the initial projected day or within a certain time window. Both algorithms were tested with actual traveling distances of municipality trucks. Traveling distances were reduced by half after optimization.

Innovation 4: Network design and tracking for sewage disposal, Mubi, Nigeria

Relevance: An existing approach that can be applied to achieve one component of the use case objective (targeting future investments in fecal sludge management e.g., where to extend existing sewer networks).

Researchers applied remote sensing and GIS to design a sewage network and aid proper sewage disposal.[72] Researchers first used this technology to assess the local topography and create a composite map of the area, which included major and minor roads, access roads, water bodies, land contours, and slope vectors. Subsequently, the researchers used composite maps (road infrastructure, gravity, etc.) to design an ideal sewage network with primary sewers (to collect wastewater from household/commercial locations), secondary sewers (to collect and transport wastewater from primary sewers), and tertiary sewers (to transport wastewater to final collection centers).

Innovation 5: Efficient garbage disposal management in metropolitan cities using vehicular ad-hoc networks (VANETs)

Relevance: An example of an existing application in a comparable sector (solid waste management) that could be applied to achieve the use case objective.

To increase efficiency of garbage collection, researchers in India simulated optimal trash collection using wireless sensor networks (WSN) and vehicular ad-hoc networks (VANETs).[73] Each residential dustbin was equipped with a light sensor that tracked how full the bin was. Once the bin was three-quarters full, a transmitter would send a signal to the nearest garbage collection vehicle, indicating the bin was ready for emptying. The vehicle would then proceed to that bin to empty it. The vehicles would also be equipped with sensors, so that if the vehicle was full, it stopped receiving bin emptying requests en route to the dumping location. This system was more efficient than sticking exclusively to a planned route. The combination of sensors and wireless communication demonstrated usefulness for optimizing routes and increasing hygiene by preventing the overflow of dustbins.

Innovation 6: Call centers for emptying jobs

Relevance: An example of an existing application that accomplishes one component of the use case objective (connecting customers with emptying services).

In Dakar, Senegal, the national sanitation agency, ONAS (Office National de l'Assainissement du Sénégal), developed the program Structuring the Fecal Sludge Market (PSMBV) for poor households to better manage fecal sludge markets and ensure access to effective, affordable sanitation facilities and mechanical emptying services.[74] PSMBV trained and certified private operators and created a call center to coordinate mechanized emptying services. The objective of the call center was to connect stakeholders (e.g., emptying operators, regulators, and households) via an improvement and optimization platform. The call center collects bids from vacuum truck operators for emptying jobs via SMS (mobile text messaging). The customer calls the call center, provides key information on their pit or tank, and confirms the date and time they want it emptied. The call center then invites emptiers to submit quotations by SMS for their service. At the end of the bidding process, the customer is notified of the lowest bid by SMS. The call center geo-references households by registering GPS coordinates of pit latrines. These points are used to help the emptier easily locate the household despite the lack of a common address system.

This scheme reportedly lowered prices by 20% and promoted increased use of vacuum trucks in low-income areas of the city. The call center conducts quality control monitoring by phone with the household and the emptier. These "switchboard" models in Dakar, Senegal, and Dhaka, Bangladesh, have shown promising results in expanding safe pit-emptying services in low-income areas where the municipal utility and/or a private entity acts as a managing group.[54,74]

Innovation 7: Improved processes for tracking pit/tank emptying jobs in Warangal, India

Relevance: An example of an existing application that accomplishes one component of the objective (scheduling emptying).

In Warangal, India, the Greater Warangal municipal corporation requires emptiers to take detailed records (household, area and location, type of septic tank, age of septic tank, date of desludging, quantity of septage, user charges collected, accidents and spillages, and the next date of scheduling) for every emptying job.[54] Emptying trucks are also fitted with GPS trackers. This information is captured through a mobile application linked to the city's property database, which is used to ensure scheduled desludging.

PARTICIPANTS

National governments (e.g., ministry of water and sanitation): National government representatives would likely be at a management level at a ministry of water and sanitation. They are familiar with data summaries and statistics and comfortable interpreting maps. They have limited time to devote to reviewing data. National government representatives would use a proposed application to:

- Download reports to inform policies and regulations. Reports could include statistics on the number of active emptying service providers, the number of emptying activities and thus the quantity of fecal sludge, the locations of fecal sludge disposal sites, average costs of emptying activities, etc.

Local governments (e.g., public health offices): Local government representatives would likely be at a management level at a public health office. They are familiar with the local context and geography and interpreting maps, as well as data summaries and statistics. They likely own a smart phone and are familiar with basic functions. They have some time to interpret and understand the data to make decisions locally. Local government representatives would use an application in the following ways:

- Track jobs to regulate locations of fecal sludge disposal and environmental contamination.
- Download reports to support local policy changes. In addition to the reports described above, this could include geographic areas not being reached by service providers (thus pointing to access/infrastructure issues or cost issues).

Service providers – public utilities: Many public utilities have a sanitation department responsible for increasing access to improved sanitation facilities and managing fecal waste.[54] Managers within these departments are familiar with data and statistics and use data to make decisions. They have enough time to understand data reports to make decisions. Public utilities may already manage databases, so there are opportunities to leverage experiences and training. Public utilities could use an application to:

- Download reports to understand if low-income communities are being served and, if not, how best to serve them.

Services providers – private (e.g., businesses or social enterprises): Decision-makers within private service providers are likely individual business operators or managers. Individual business operators and truck drivers might have limited experience understanding data, although they are likely familiar with smart phone applications and basic map software. Managers are likely very familiar with data, data analysis, statistics, and mapping software, and can easily navigate mobile applications. Both types of service providers have enough time to understand the application and use it throughout the day. Private service providers would use an application in the following ways:

- Receive job requests and determine when to visit which pit latrines/septic tanks.
- Decide where to dispose of fecal sludge from an emptying job (depending on current location and traffic).
- Decide to whom to market services, using an emptying frequency schedule tracked by the app. Service

providers could push notifications to prior customers to inquire about their emptying needs.

Customers: Customers are local residents. They likely own a phone or borrow a family member's phone and are familiar with basic functions. Some customers have access to a smart phone and will be able to download the mobile app; however, some customers are limited to a simpler phone (e.g., flip phone, feature phone). An application would have to be adapted to this type of phone to allow customers to call/text the system to scheduled services. Customers have enough time to learn to use the app. They would use it in the following ways:

- Review service provider ratings and availability
- Schedule upcoming pit emptying activities
- Track the real-time status of an emptying event (e.g., where the truck/manual emptier is, where they are taking the fecal sludge)
- Access records of their previous emptying events (cost, which service provider they used)
- Review and provide feedback on the emptying event

Developer: A developer is an individual (or team) that develops mobile applications. The developer has advanced computer programming and design skills. They are familiar with phone applications, how they function, and how to iterate the application development to suit participant needs. Their time would be devoted to the development of the application during the building phase, and maintenance and upgrades thereafter.

OUTPUTS

- A mobile phone application developed for use by emptying service providers and customers.
- An online web application synced to the phone application to serve as a data dashboard. Users would be able to log into the desktop/online version of the application and download standard and customizable reports.

WORKFLOW

Step 1: Identify a location to develop and test the application

The team needs to identify a location to develop and test the application. The location for initial development and piloting should be a dense urban area with some areas of low-income housing, to verify that these markets can be included in the application. The location should preferably have at least one legally established fecal sludge dumping location. This would help the local government track its use and determine if additional legal dumping areas are needed. In the test location, at least one participant passionate about the optimization of emptying activities should serve as a community champion. To generate further demand for the application, the implementer should promote it through household visits, posters, and opt-in SMS reminders.

Step 2: Gather existing locally available data and regulatory requirements

The local and national government offices in the location should provide the developer with existing geo-referenced information on the locations of low-income areas, road maps, administrative boundaries, water tables, and existing dumping locations. The sanitation utilities in the location should provide the developer with existing geo-referenced information on the extent of sewer networks (if applicable).

Step 3: Develop a mock-up of the application features and flow

The developer would start by designing a basic application with relevant features and flows. They would request initial feedback from a sample of each stakeholder group (local governments, national governments, service providers, customers), to understand their needs. The developer would then use the feedback to create a preliminary application.

Step 4: Application design iteration (the "pilot" phase)

The pilot phase tests operation of the application at a small-scale, while training key staff and collecting information from customers and staff on how to improve the application. The developer would try to understand if unnecessary features can be removed to increase simplicity and usability. Without an appropriate design and relative ease of use, the application will not likely be adopted by all participants. Iteration should continue until achieving general approval from representative stakeholders.

Step 5: Launch application

Launching mobile application requires accounts with app stores. The developer could create a desktop/online dashboard to provide analytics and report outputs. Users should have access to standard and customizable reports (examples under participants section). Training should be held with local governments, national governments, and service providers to increase familiarity and spread word of the application. The developer should work with local governments, national governments, and service providers to advertise through social media channels, press releases, and word of mouth.

Step 6: Continuous support (the "scaling" stage)

Lessons learned and user feedback during steps 1 and 4 should be considered to improve the applications services and to prepare for large-scale operation. Each type of user should only be able to access the relevant components of the application. For example, customers should not have access to reports that are being generated for government users. New data (e.g., new roads, new dumping locations) could be added as it becomes available. Additional features could be added as needed.

Step 7: End-user interactions

Utilities and private service providers would need to download the mobile application to use it on their phones. They would require a smart phone with location tracking abilities (GPS). They would need to log in to receive job requests and set an emptying schedule. When completing a job, they would answer a few basic questions about the emptying activity. Once they emptied the designated pits/tanks and filled their truck, the application would suggest an optimal route for disposing the fecal sludge.

When utilities and private service providers are not actively completing a job, the application could suggest customers who might need services. They might push an offer for their services directly through the application. Utilities and private service providers would also be able to log into the desktop/online dashboard. They can download standard or custom reports to help with their business needs.

Customers wouldneed to download the mobile application to use it on their smart phone. They would have to log into the application (or can select to stay logged in) to schedule an emptying service. They would select their preferred service provider based on availability and ratings. During disposal, they would track the truck to ensure appropriate disposal (or an estimated return time if more than one trip is needed). At the end of an emptying event, they would rate the service provider on their job. The application could send notifications of when their pit might be due for another emptying.

DATA SOURCES

Existing data:

- In most urban contexts where this application could be applied, there is likely existing GIS data collected and maintained by local governments (specifically roads, water bodies, possibly low-income areas).
- For some urban areas, geographic coordinates for existing, legal fecal sludge disposal sites are available, or professional knowledge of these locations can support collection of the geographic coordinates.
- Often service providers or local government officials already have familiarity with neighborhoods and sanitation access issues. This local knowledge could be translated into data to be used within the application.
- Existing software such as Google Maps is increasingly used in urban areas in low- and middle-income countries. Google Maps could be the foundation for the map and tracking component of the application.

Data gaps:

- Existing maps do not often give the best picture in intricate, high-density urban areas. In these areas, although there may be a road, it might be impassable by vacuum trucks due to the road width, road condition (muddy after a rain), or road closure. This data would need to be captured for development of an application.
- Traffic in many urban areas can be debilitating. Often routes are poorly mapped due to unknown traffic conditions and abovementioned road condition issues. Ideally, traffic conditions would be incorporated into the application to allow for optimal route determination.
- Service providers do not track household/business emptying needs unless they have standing emptying activities (such as a weekly emptying for a large office). The application would be constantly recording emptying activities, which would allow service providers to track possible upcoming jobs, predict busier times, and target marketing activities.
- GPS coordinates are often not available in dense, urban areas, so one opportunity to address this gap is to include a GPS recorder in the application.

BARRIERS

Scale: For the application to be successful, it should be used at a large scale (i.e., within the jurisdiction of a single town, city, county, or service provider) to maximize its effectiveness. It may be difficult to incorporate all moving pieces at that scale.

Buy-in: Potential challenges may include buy-in from emptying service providers and regulators. There may be concern in terms of increased oversight of their work that requires developing a new business model and offering services during the transition. In addition, consumer buy-in may require community sensitization and application piloting.

Data literacy: There may be challenges with data literacy. To address this, training on application use and functionalities should be provided before the pilot and launch.

Data gaps: There are large data gaps in urban area mapping, particularly among low-income areas that may not be geo-referenced. This can make it difficult to employ the application given the uncertainty of area boundaries or numbers of households. One way to mitigate this uncertainty is to use satellite imagery to determine the population density of an area and other environmental factors.

4. SANITATION: UNDERSTANDING SANITATION COSTS

OBJECTIVE

To provide sub-national governments, service providers, funders, and research organizations with a software application to compare the costs of sanitation interventions and promote data-driven decisions.

DESCRIPTION

The SDG call for universal access to sanitation is not on track to be met by 2030.[1] A central challenge to meeting the SDG targets is selecting the most appropriate sanitation option(s) for each local context.[75] Providing 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.[75,76] Low quality or limited cost data forces assumptions and excludes cost categories, which leads to "incomplete and potentially misleading results."[77] Trémolet et al. note that, "despite decades of field experience, reliable estimates for the hardware and software costs of sanitation access are still scarce."[78]

Previous studies of sanitation economics lacked the costing data necessary to cover life-cycle costs over the full sanitation chain from construction to disposal and reuse.[75,76] Another typical study limitation is including only one component of the sanitation system or only the first year of operation in reported costs.[79] Manga et al. found that very few urban local governments or water utilities calculate the real costs of managing onsite systems, which should include the costs of emptying sludge, transporting it for treatment, and operating treatment facilities.[80] While recent initiatives seek to develop life-cycle costing applications,[81,82] some of which are comparative,[83,84] they are all designed for citywide solutions. There may be opportunities to use high-resolution contextual data to instead customize sanitation services at the neighborhood level, thereby increasing the suitability, cost-effectiveness, and sustainability of services. Possible future technological developments might include compiling widespread design data for key sanitation options into typical (or even standardized) bills of quantities (BoQs; lists of materials and services required for implementation). Such a application could be used by planners to obtain substantiated cost estimates by adding context-specific adaptations (e.g., of the types and quantities of materials to be used) and entering local unit prices.

Eawag's Compendium of Sanitation Systems and Technologies describes numerous sanitation options.[85] Some common options that could be considered when designing a proposed application include:

Offsite sanitation systems[86]

- **Conventional sewers** – Pressurized sewers rely upon pump stations and are appropriate in flat topography; in rocky, hilly, or densely populated areas; or in areas with a high groundwater table where deep excavation is difficult. Gravity-based sewerage may be preferable in densely populated areas with a reliable supply of piped water, local topography that allows for gravity-flow sewers, and/or in areas immediately adjacent to middle/high income neighborhoods where a sewer system is already present.

- **Simplified sewer systems** (i.e., condominial sewers[85,87]) – These smaller, shallower sewer networks connecting multiple dwellings or office buildings are most applicable and cost effective in areas with rocky ground or a high groundwater table and a reliable piped water supply. The systems can be installed in all types of settlements and are more appropriate in dense urban areas where space for onsite technologies is limited.

Onsite sanitation systems[88]

- **Pit latrines, septic tanks, and compost latrines** are cost-effective in stable and permeable soils with low groundwater levels, and low- to medium-density residential areas. Accessibility via vacuum trucks or carts promotes easier sludge removal.
- **Composting toilets** are feasible in areas under the most difficult (e.g., rocky) soil and high groundwater conditions.
- **Vault toilets** are low-volume flush toilets that discharge into a sealed tank or vault. They require regular emptying, but are cost-effective in high-density settlements with a high water table. Accessibility via vacuum trucks or carts promotes easier sludge removal.
- **Container-based sanitation** is a service that provides portable, self-contained toilets to households and collects the accumulated wastes every few days. It is cost-effective in areas where high customer penetration can be achieved, for example high-density tenant housing with no sanitation facilities on the premises.

DECISION STATUS QUO

Decision-makers typically select sanitation service options without considering cost-benefit accounting for the full sanitation chain (e.g., long-term operation and maintenance costs of sludge treatment, reuse, and/or disposal).[89] In addition, implementers often select sanitation interventions using only the lowest dollar amount needed on a city-wide level (capital investment required from the public budget) or historical political choices.[89] Costs are generally calculated for the entire city population and reported per capita, which is potentially less accurate than focusing on the specific needs of those still lacking sanitation access. **Without a detailed understanding of the specific costs of sanitation interventions, governments and implementors struggle to select the most cost-effective approach for optimizing sanitation solutions.**

Interviewees indicated that typical decision practices lack key cost information and are more prone to failure.[90]

- "Most initiatives driven by development partners or stakeholders fail because they don't understand the beneficiaries' needs and the processes are not people-centered/driven to help understand cost implications. Sanitation components are hurriedly taken over by governments and right approaches to drive projects lose sight of the social fabrics of the targeted community thus depicting stakeholders' lack of this kind of data."
- "Governments without sanitation programs are not interested to understand cost[s] of providing safe sanitation and those running sanitation programs are intrigued when they get a share of the money provided for the project."
- "Political interest hampers urban sanitation decision-making process[es] as well as access to costing data, in terms of who should take part in the process, and most of the data users do not understand type of decisions to make from the data."

DEMAND

The ability to anticipate the full economic costs of sanitation projects would help avoid project failure and promote sustained use. Cost data found in the literature do not always cover the full sanitation chain because existing systems in low- and middle-income countries are often incomplete and/or cost data simply does not exist.[75] In addition, data on life-cycle costs of sanitation solutions is scarce,[75] as are reliable estimates for both the hardware (infrastructure) and software (behavior change and sensitization) costs of improving sanitation access.[78] Reported data sometimes do not distinguish clearly between urban and rural areas, sufficiently specify sanitation options, provide breakdowns of overall cost, or make use of multiple metrics (e.g., cost per capita vs. unit cost).[75]

The limited availability of sanitation cost data remains a key constraint for practitioners aiming to develop reliable cost estimates and budgets to construct and operate sanitation and wastewater infrastructure. In addition, data are often very specific to one project or context, and material and labor costs are not documented in detail (e.g., lump sum aggregate rather than unit rates and quantities for individual components), making it difficult to analyze the various costs. BoQs provided by third-party organizations lack the corresponding metadata (design information drawings and pictures).[91]

Interviewees elaborated:

- "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."[89]
- "Engineers and project managers struggle to provide site-specific cost estimates that allow for sound technology decision-making and budget planning in sanitation programming. Estimating capital and operational infrastructure costs is not easy, especially considering all the context-specific and variable factors that determine the total costs."[91]
- "Sanergy [a private sanitation company] would like to understand household's current sanitation options and their costs (e.g., pit latrine, toilets that drain into the open, open defecation, etc.) – this [is] useful data for Sanergy because they can compare a household's current sanitation costs with what they are offering. If a toilet that drains into a river is cheaper than Sanergy's offer or free, Sanergy would like to know to what extent money is a constraint for joining Sanergy's program, even if the household acknowledges the hygiene, privacy, and other benefits of the proposed toilet."[66]

OTHER DATA APPLICATIONS

Outputs from a proposed use case application(s) could be used for other decision-making applications as well. Utilities and service providers could use outputs to structure user fees, design tariffs and subsidies, and improve subsidy allocation. Implementing organizations could use outputs from the application to designate zones within a city where sanitation access is low, which is likely to correspond to low-income areas. Sanitation zone delineations could be used to support development projects related to drinking water, education, housing, and health.

EXISTING AND UPCOMING INNOVATIONS

Innovation 1: International Rescue Committee (IRC) WASHCost life-cycle costing tools, Ethiopia

Relevance: Tool for collecting data for a life-cycle cost approach to water and sanitation service delivery at the district level. It is an existing tool that could be further developed to achieve the use case objective.

IRC developed life-cycle cost guidance at the district level in Ethiopia to collect multiple forms of data that underlie a tool to improve water service delivery.[92] This approach could potentially be applied for sanitation service delivery. The guide consists of two steps:

1. Collect data on infrastructure components, source type, functionality, age, management, water quality, and water reliability of all water schemes. The data are used as a basis for making decisions on rehabilitation and maintenance.
2. Establish the cost of current water service delivery using data from government water offices at different levels: local, district, municipal, zonal, regional, households, or water user committees, depending on the administrative system.

Collected data is entered in a Microsoft Excel format and uploaded to the mWater application,[93] which then estimates the gap between existing services and full coverage at the desired service level. The application also produces data to characterize the affordability and adequacy of tariffs paid by households.[94]

Innovation 2: CLARA Simplified Planning Tool (SPT), Ethiopia (IRC)

Relevance: Tool for local planners to use to compare different water and sanitation systems during the early planning phase of a program. It is an existing application that can be applied to achieve the use case objective.

CLARA SPT is a software application that calculates and compares full costs (investment, operation and maintenance) of water and sanitation systems at early planning phases.[95] Sanitation solutions implemented within the SPT are grouped by their functionality. Thus, each sanitation solution has a defined standard design selected from several sizes, and costs are calculated from BoQs. BoQs assumptions include a short description of the technology, design, cost range, assumed lifespan, and operation and maintenance costs. The design has been adapted to local contexts in African countries such as Kenya, South Africa, and Ethiopia.

Innovation 3: Economics of Sanitation Initiative (ESI) Toolkit

Relevance: A tool developed by the World Bank to generate evidence on the economic benefits of improved sanitation. This existing tool can be applied to achieve elements of the use case objective.

The ESI Toolkit enables stakeholders to generate analyses that describe improved sanitation costs.[96] It uses algorithms to calculate and compare results of various projected solutions across different time periods and areas. Its cost-benefit and cost-effectiveness modules combine cost and benefit results to produce standard indicators demonstrating socioeconomic return on sanitation investment. Results are easily shared among users.

Innovation 4: Community-Led Urban Environmental Sanitation (CLUES) toolbox – Sanitation costing tool

Relevance: Existing example of the use case objective

The CLUES toolbox is an automated, BoQ-based, Microsoft Excel tool that estimates construction and maintenance costs of select sanitation technologies.[97] It assesses the cost implications of required material, labor, and maintenance for a range of sanitation technologies. The user interface differentiates between basic (low-cost) and advanced (higher-cost) options. Cost estimates are calculated from generalized BoQs, local unit rates for materials (e.g., the price of one bag of cement), and local rates for skilled/unskilled labor.

This tool does not estimate total life-cycle costs and generalized BoQs are not available for all sanitation technologies because some have highly variable design alternatives depending on local conditions, preferences, and standards.[91,97]

Innovation 5: WASH SDG Costing Tool, World Bank & UNICEF

Relevance: A tool used to calculate the investments and financing gap needed to fulfill the SDG targets (6.1 and 6.2) at a country level. It is an existing tool that can be applied to achieve the use case objective.

The WASH SDG Costing Tool uses a model developed in Excel to calculate the population that can be covered by each type and level of sanitation service, given expected population changes between 2015–2030.[98] It also allows customization of the baseline year. Capital investment costs are clearly distinguished in the model, and software costs are included separately under the capital costs. Service levels, coverage numbers, and cost estimates are disaggregated for urban vs. rural areas. With further adaptations, this tool could be used to make sub-national estimates (e.g., province, region, or state). Despite validation of country-specific values, the outputs should be interpreted with caution, because the cost model is fairly basic with uncertainties remaining for many of the underlying values.

Innovation 6: Climate and Costs in Urban Sanitation (CACTUS), University of Leeds

Relevance: The CACTUS tool provides the cost of functional sanitation infrastructures across the "sanitation value chain" (from capture to safe reuse or disposal).[99] It is an existing tool that can be applied to achieve the use case objective.

The CACTUS tool describes costs of different sanitation service delivery options as total annualized cost per household and total annualized cost per capita. Total annualized costs include capital costs annualized over the lifetime of the relevant infrastructure or equipment, costs of capital and discounting, and the annual operational costs associated with that element of the sanitation service delivery option. All costs are normalized in a database to a single currency and date to allow comparisons.

PARTICIPANTS

Sub-national governments (e.g., district governments, public health offices): Local government representatives tasked with implementing sanitation programming and setting community targets would likely be at a management level. They are familiar with the local sanitation context and are comfortable managing and interpreting data. They have basic technical skills, but will require some additional tool-specific training. Sub-national government representatives could use applications to:

- Run scenarios before deciding on upgrade plans for a city.
- Prioritize funding allocations based on selected intervention costs.

Service providers – municipal utilities: Municipal utilities often have a sanitation department responsible for increasing access to improved sanitation facilities and managing fecal waste.[148] Utilities typically bear some costs related to fecal sludge management (FSM) and treatment/disposal. Utility managers are familiar with data and statistics and regularly use data to make decisions. Municipal utility managers may already manage databases, so there are opportunities to leverage experience and training. Managers will have some technical skills, but some staff will require training to use applications in addition to written guidance. They will have some time available to use applications. Utility managers could use applications to:

- Run scenarios before deciding on upgrade plans for their distribution networks.

Service providers – private (e.g., businesses or social enterprises): Decision-makers within private service providers are likely individual business operators or are at a management level within a business. Managers within enterprises are likely familiar with data, data analysis, statistics, and mapping, and can easily navigate software. Private service providers will likely have the technical understanding necessary to use applications independently. They will have some time available to use applications. Private service providers could use applications to:

- Run scenarios to evaluate where their innovation is more efficient to help target customers or make business adaptations.

Funders: The funder most likely represents a grants manager for a large foundation or a contract project manager for a large multilateral funder. Funding often has strings attached and is earmarked for a particular solution; thus, it is important for donors to also understand the optimal solutions. We expect donors to have the access to technical understanding necessary to use applications independently. However, they might not be familiar with all of the technologies presented and might need some additional guidance. They will have sufficient time to use applications to allocate funds. Funders could use applications to:

- Suggest or prioritize sanitation needs within a grant/contract.
- Understand the cost requirements to successfully implement a sanitation strategy.

Universities and research organizations: Universities and research institutes often advise governments and utilities on appropriate solutions. They are familiar with managing and interpreting data, though they will be less familiar with local contexts. Researchers have the technical understanding necessary to use applications independently. They can use applications to help advise service providers and governments. If research involves deploying an evidence-based sanitation innovation, applications could assist in locating target areas for testing.

Data scientists: A data scientist's primary expertise is in developing models, conducting analyses, and providing data interpretations. Their time would be devoted to the development of applications during the building phase. A data scientist would need to be minimally involved on an ongoing basis to update costs, features, sanitation options, etc.

OUTPUTS

- A database for index values would first be developed for various sanitation-related costs specific to a city. The goal of the database is for users to reference values for similar cities to fill data gaps.
- The application itself will be an online application which will perform two primary functions: 1) mix data derived from secondary datasets, satellite imagery, and user-input location-specific data to define neighborhood zones and suitability for differing interventions within each zone, and 2) calculate city-wide sanitation costs using the mix of neighborhood-level interventions. The costing calculations would build upon those in previously developed Excel tools. If the user is missing any required data, reference values could be pulled from the costing database described above. The application could compare predicted costs for each sanitation option. The user could select the stages of the sanitation value chain to include in the overall cost comparison (as necessary, due to possible data gaps).
- Users should be able to download data and maps.

WORKFLOW

Step 1: Desk review

The development team would perform a desk review of existing Excel life-cycle costing tools for sanitation interventions (e.g., sewerage, septic tanks, improved pit latrines, and container-based sanitation). Examples include: WASH SDG Costing Tool (World Bank & UNICEF), CLARA SPT (WSP), SanCost (Aquaya), and the CACTUS Costing Tool (University of Leeds).

Step 2: Evaluation existing applications

If any existing tools are adequate, it would be used for the cost comparison in Step 7. We propose a goal not to recreate an existing tool, but instead to create a application that addresses the various mixes of interventions. This could mean adapting an existing tool and embedding it within the new app. Alternatively, a custom version could be created by combining features from two or more existing tools.

Step 3: Select pilot

The team should identify a city in which to pilot the app.

Step 4: Define geographic scope

The applications should define zones where the need for sanitation access is likely to be high. If available, high-resolution income or poverty data can be leveraged to define target zones. If no data exists, the team can map areas manually using of neighborhood boundaries and stakeholder interviews.

Step 5: Select secondary datasets

Secondary datasets will be identified for situation analysis and data treatment in the pilot city. Target variables could include population density, topography, distance to treatment plant, and the water table level.

Step 6: Define "rules"

Developers can define the "rules"/calculations for defining suitability for each intervention within each zone. Inputs should include, but are not limited to, the type of technology, labor costs, material and utility costs.

Step 7: Develop calculations

The next step is to set up calculations for combining costing data with suitability to determine the optimal combination of sanitation interventions and the population(s) served by each. Costing calculations (derived from existing tools) will be expanded to include overall city costing using the suggested combination of neighborhood-level interventions. Costs for achieving universal sanitation can then be determined.

Step 8: Piloting

The development team will need to test the application in pilot cities, incorporate feedback from participants, and iterate the application until satisfactory.

Step 9: Build web application

A web application could be created to host the software, allowing users to upload costing data, and produce outputs describing optimal city-wide sanitation plans. For example, the application could be linked to the World Bank sanitation costing database.

Step 10: End-user interactions

Users will require a computer or smartphone with internet access to access the web-based application and to specify the geographies of the urban areas and low-income zones for which interventions are needed. Users can choose to add their data to the database. A cost model section will initially establish cost predictions for each sanitation intervention option. Then, running pixel-by-pixel suitability calculations would produce a map of optimal sanitation solutions across the selected geographies. Cost data will be pulled to determine the costs associated with the optimal combination of interventions. The model will result in multiple alternatives where appropriate.

DATA SOURCES

Existing data:

- Daudey compiled the main determinants of urban sanitation cost from numerous peer-reviewed and grey literature.[75] The following data could be included in the costing tool: type of technology, labor cost, material and utility cost, energy cost, end use of treatment products,. Other available data include: density, topography, level of service provided by the sanitation system, soil condition, population served by system, distance to treatment facility, climate, business models, and water table level.
- Existing costing models contain data on typical design schemes that can be imported into the proposed app. The user would be able to customize these variables, if desired. Given the localized nature of cost data, no universal values are available for construction materials, emptying services, treatment, and disposal. Users of the application are expected to have some knowledge of the local context, but for data gaps, the application could link to a database of reference values maintained by the World Bank.
- The sanitation suitability application will draw on secondary datasets related to population density, topography, soil condition, climate, and water table level, all in the format of digitized rasters:
    - Population density (100 x 100 m) sourced from WorldPop.[100]
    - Topography/Digital Elevation Model (1-m interval) sourced from Google Maps/Earth or Shuttle Radar Topography Mission data[101] available from the US National Aeronautics and Space Administration (NASA), NGA, or USGS.
    - Soil type (250 m) sourced from the International Soil Reference and Information Centre (ISRIC).[102]
    - Water table height (5 km) sourced from the British Geological Service.[103]

Data gaps:

- Higher-resolution water table information, which might be available locally, would influence waste containment options and costs.
- Understanding how to use soil map information requires detailed knowledge of all soil types and those most influential for sanitation (e.g., which are prone to erosion and/or flooding).
- Traffic can greatly affect the feasibility of sludge transportation to the treatment facility. Incorporating traffic information will be challenging as it varies widely by space and time.
- Width of roadways is important for determining whether vacuum trucks are able to enter an area. Population density could be used as a proxy.

- A method needs to be developed for determining proximity to fecal sludge treatment facilities. Satellite imagery and/or OpenStreetMap can be used to establish the location of the treatment facility either manually or via machine learning.

BARRIERS

Scale of data: Since there is a need for highly localized data, particularly for sewage treatment and disposal, this approach is more difficult to apply across multiple cities and countries. It will be possible to develop a database in tandem with a applications that users have options for reference values for various components.

Assumptions: "Rules" to determine sanitation suitability will include assumptions that may not be appropriate in all contexts.

Regular updating: Urban areas are prone to rapid development, requiring continued maintenance to update secondary datasets (e.g., for population density and roadways). Additionally, treatment costs may change as urban areas grow and change.

Buy-in: Potential challenges may include buy-in from service providers and regulators.

Data literacy: There may be challenges with data literacy among users. To address this, training on application use and functionalities should be provided before the pilot and the launch.

Complexity: It may be challenging to consider social or cultural influences as variables within the application. Optimizing intervention recommendations requires that they fulfill legal requirements, benefit the environment and public health, and are socially appropriate.

5. HEALTH: ANTICIPATING WATERBORNE DISEASE OUTBREAKS

OBJECTIVE

To provide public health authorities with trajectories of waterborne disease outbreaks to guide targeted prevention and mitigation efforts.

DESCRIPTION

The International Health Regulations (2005) are a legally binding agreement among WHO member states to proactively protect global health security. They obligate countries to detect, assess, and report public health events such as outbreaks. While the Regulations are not disease specific, they provide a framework for global disease surveillance and reporting. Several waterborne diseases are commonly reported.[104]

Waterborne diseases are infections acquired through consumption of, or contact with, pathogen-contaminated water. There are many waterborne diseases such as those caused by viruses (e.g., rotavirus, Hepatitis E), bacteria (e.g., cholera, typhoid), protozoa (e.g., cryptosporidium, giardia), and helminths (e.g., schistosomiasis, Guinea worm). Collectively, waterborne pathogens infect millions of people every year, resulting in a considerable health burden, predominantly through diarrheal disease in low-income countries among children under five.[105]

Waterborne pathogens can persist in environmental reservoirs. Disease propagation is primarily through the fecal-oral route and attributable to unsafe water, sanitation, and hygiene (WASH) practices.[105] Waterborne disease outbreaks can therefore largely be prevented by widespread use of well-maintained WASH infrastructure, programs, and practices. Though substantial progress has been made in expanding coverage of safe water and sanitation services, billions of people globally remain without access.[106] In low-income countries, even where safe water and sanitation services exist, they are not always used or delivered consistently. For example, rural water users often rely on both improved and unimproved sources.[107–109] Further, as urbanization continues to crowd cities in the Global South, already stressed WASH services may struggle to keep pace.[110,111] These realities leave large populations highly vulnerable to waterborne disease outbreaks; to compound the issue, natural disasters, conflicts, and climate change exacerbate the risk of outbreaks.

Although the long-term solution to minimizing waterborne outbreaks is achieving universal WASH coverage, a long, challenging path lies ahead. Medical countermeasures (e.g., vaccines, antibiotics) have and will continue to play an important role 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 service delivery interventions.[112–117] While the connections between environmental and hydro-climatological conditions and disease occurrence have long been recognized, in recent years, researchers have advanced understanding of these relationships and refined capabilities to project waterborne disease risk spatially and temporally.[118–120] With forecasted risk maps for waterborne diseases, public health professionals could more effectively and efficiently allocate resources for outbreak prevention and control.

DECISION STATUS QUO

The predominant approach to disease surveillance focuses on retrospective identification and response. 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. Confirming an outbreak depends partly on endemic disease levels; in some situations, one confirmed case is enough to trigger action. In other cases, several confirmed cases might be needed to prompt action. In low-resource settings, risks are higher due to fewer clinical laboratories and limited pathogen testing capacity. For many infectious diseases, propagation within the population has often already started by the time cases are reported and confirmed by health authorities. This leaves little-to-no lead time for responders. Unprepared healthcare systems then react slowly to implement measures that could curtail further spreading. **Without anticipating scenarios of waterborne disease outbreaks, public health experts and governments cannot prepare targeted prevention and mitigation efforts to reduce the global burden of disease.**

DEMAND

The imperative to reduce the negative health and economic impacts caused by waterborne diseases is clearly defined within the framework of the SDGs, explicitly by SDG 3 ("Ensure healthy lives and promote well-being for all at all ages") and implicitly by several others. To that end, 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."[121,122] Several academic groups (e.g., Jutla Research Group, University of Florida; InForMid, Tufts University; Alexander Research Group, Virginia Tech) are working on developing applications for anticipating waterborne disease outbreaks. Practitioners also show interest in disease forecasting, as 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."[123]

OTHER DATA APPLICATIONS

Apps used to predict waterborne diseases will be applicable to other diseases, most notably those with climate and/or environmental drivers, such as mosquito-borne diseases.

EXISTING AND UPCOMING INNOVATIONS

Innovation 1: Software applications and platforms for vector-borne disease (e.g., malaria, dengue, chikungunya) monitoring, mapping, and forecasting

Relevance: The disease transmission process is not exactly the same as for most waterborne diseases, but there are similarities, especially seasonality. These existing applications can contribute to achieving the use case objective.

A few global- and national-level examples include:

- **VectorMap**: "a web-based platform providing access and visualization of global vector distribution data."[124]
- **CHIKRisk:** an online platform that provides climate-based risk maps for chikungunya fever. With the online platform, users can visualize current and forecasted risk on a global scale.[125]
- **Epidemic Prognosis Incorporating Disease and Environmental Monitoring for Integrated Assessment (EPIDEMIA)**: an early warning system which supports malaria forecasting in epidemic-prone regions of Ethiopia.[126]
- **Artificial Intelligence in Medical Epidemiology (AIME)**: a startup aiming "to become the global reference center for disease and epidemic predictions." While yet to be validated on a large scale, the system they developed for dengue prediction shows promise, for guiding outbreak control interventions.[127]

Innovation 2: Disease and pathogen distribution maps

Relevance: Anticipating future risks requires an understanding of past and current risks, which may in part be influenced by disease distribution. Existing maps can contribute to achieving the use case objective.

Disease and pathogen distribution maps are available at the global (e.g., Global Water Pathogen Project), regional (e.g., HEALTHY FUTURES), and national (e.g., Global Atlas of Helminth Infections) levels.[128-130] These could provide useful examples for developing waterborne disease risk maps.

Innovation 3: Famine Early Warning System Network (FEWS NET)

Relevance: FEWS NET applies a "prediction and prevention" paradigm to food security. This an existing technology can contribute to achieving the use case objective.

FEWS NET is a USAID-funded and managed collaboration of scientists and implementers providing early warning and analyses on food insecurity.[131]

PARTICIPANTS

National governments (e.g., ministry of health; regional, county, or district public health offices): National government representatives would likely be at a management level at a ministry of health. They are familiar with data summaries and statistics and comfortable interpreting maps. They understand disease prevalence metrics and may have some understanding of transmission pathways; however, they may be less skilled at interpreting risk and understanding complex climate-environment-disease relationships. They have limited time to devote to reviewing data. National government representatives could use forecasted disease risk maps in the following ways:

- Develop policies to support interventions that address the greatest public health and economic concerns.
- Allocate resources quickly to the areas of greatest need.

Local governments (e.g., regional offices of the ministry of health): Local government representatives would likely be at a management level at a local health ministry office. They are familiar with local context and geography and data summaries and are comfortable interpreting maps. They understand local disease prevalence and may have a basic understanding of transmission pathways; however, they will have little to no skill at interpreting risk and understanding complex climate-environment-disease relationships. They have a moderate amount of time to devote to reviewing data. Local government representatives could use forecasted disease risk maps in the following ways:

- Plan interventions to addresses the greatest public health concerns.
- Allocate local resources to the areas of greatest need.
- Work with vulnerable communities to minimize risks.

Multilateral public health agencies (e.g., World Health Organization) and additional implementing partners (e.g., CSOs): Stakeholders at multilateral public health agencies and CSOs would likely be at a management level. They are very familiar with data summaries and statistics, and have a basic understanding of disease transmission. They have varying technical capacities to interpret risks and understand underlying drivers of disease transmission. These stakeholders could use forecasted disease risk maps to:

- Prioritize when and where to intervene, whether by providing monetary, material, and/or personnel support.

Data scientists: A data scientist's primary expertise is in developing models, conducting data analyses, and providing data interpretations. For disease trajectories to remain relevant, the data scientist will need to be engaged on an ongoing basis. Data scientists conduct the "behind the scenes" work to ensure that decision-makers have up-to-date information.

OUTPUTS

- Interpreted disease risk **maps** should be shared with relevant decision-makers on a monthly or quarterly basis.
- "Risk reports" could be housed on an open **online platform**, or only accessible to relevant parties.
- Curated **datasets** of disease model inputs could also be housed on an online platform. This would likely be most relevant to researchers.

WORKFLOW

Step 1: Model environment-disease transmission relationships

The team needs to first conduct a literature review to better understand environmental and climatic drivers and transmission models of targeted waterborne diseases. They can gather data from existing data sources, covering each identified environmental and climatic driver and historical disease transmission rates.

Step 2: Develop exploratory analysis

Data scientists should complete exploratory spatiotemporal statistical analyses to explain disease occurrence in space and time and identify risk factors.

Step 3: Pilot location selection

It is important to pilot the mapping exercise in one location. The team should identify a country to develop a location-specific model. Local health offices must be interested in providing relevant data and interested in using the outputs.

Step 4: Map disease risks and vulnerabilities

The team should validate risk factors using the available historical data and develop a model that predicts future risk. The team will share initial outputs with stakeholders and incorporate feedback as necessary.

DATA SOURCES

Existing data:

- Disease data reported through national surveillance systems (e.g., ministry of health databases) may be subject to quality and quantity issues. Reported case data are likely to be available upon request but relatively limited compared to high-income countries.
- Other disease databases include the Global Enteric Multicenter Study (GEMS)[132] and Global Infectious Diseases and Epidemiology Online Network (GIDEON).[133] These may be useful for exploring climate-environment-disease relationships to develop explanatory risk models.
- Completeness and accuracy of water supply and sanitation infrastructure databases may vary by country. Access to WASH services is an important factor in risk modeling. National georeferenced water and sanitation infrastructure databases may be available upon request from appropriate ministries. National household surveys, such as DHS, could be useful for obtaining sub-national WASH coverage data.
- Roads and public health infrastructure (e.g., healthcare facilities) can be mapped through publicly available GIS databases. These data will contribute to the understanding of how proximity to roads, urban areas, and health infrastructure influence transmission risk.
- Climate data (e.g., precipitation, temperature) observed at weather stations may be available from national meteorological agencies. Data quantity and quality may be limited, but where sufficient, incorporating observational data could improve model outputs. Climate data can also be derived from open access satellite and remote sensing sources. Transmission of many waterborne diseases is influenced by climate, so understanding these links is important for mapping current and future risks.
- Land cover, land use, terrain, and surface water bodies can be obtained from satellite images and remote sensing, with several open access options.[134] Understanding relationships between environmental factors and disease would allow mapping over larger geographic areas.
- Population density and poverty data may be available at low or high resolution.

Data gaps:

- Modeling microbial outbreak risks relies on pathogen-specific data on occurrence, fate and transport, transmission, and removal by water treatment processes, which is often limited or unreliable.[135] When actual data are unavailable, researchers are often left to use assumptions or "worst-case-scenario" values.
- Uncertainty remains around the health impacts of a pathogen at various levels of exposure, particularly low doses.[136] Researchers try to fit each pathogen to a statistical distribution to evaluate the "dose-response" relationship, but this may stem from limited study. Further, a dose-response relationship may not be available for all strains of an organism, or understood for subpopulations of the community that have greater susceptibility.
- Modeling multiple exposure pathways adds complexity and can reduce reliability of risk models, particularly when trying to understand immunity and infectivity.[136]
- Clinical surveillance data that tracks disease occurrence and spread is notoriously unreliable and limited in some locations. The speed at which this data becomes available to the public can limit the effectiveness of the model, particularly where digital reporting is uncommon. Data can also be aggregated during reporting, and language and cultural barriers can affect data literacy and harmonization (for example, where patient interviews are used).[137]

BARRIERS

Data completeness: Most countries have reporting requirements for many diseases, so epidemiological data exist, but health records are notoriously incomplete and have some degree of inaccuracy.

Digitization: Until recently, paper-based medical records were the norm in most of the world and still are in many contexts. Digitizing and standardizing historic records would provide useful data, but this would be an enormous undertaking. With the proliferation of mobile monitoring and reporting systems and internet-connected point-of-care devices, health data should become increasingly available, even in low-resource contexts.[138]

Communicating model uncertainty: Decision-makers with limited data literacy need to be able to understand and meaningfully interpret uncertainty. To the extent possible, involving decision-makers in model development and providing training opportunities would be beneficial.

6. COMMUNITIES: INTERPOLATING HOUSEHOLD DATA

OBJECTIVE

To provide governments, implementers, and researchers with improved and comprehensive household-level data (e.g., maps of high-resolution WASH indicators of interest) by interpolating actual data gathered through household surveys by governments, implementing organizations, or researchers.

Decision: Stakeholders could use household-level data maps to tailor decision-making and implementation solutions to a location's context.

DESCRIPTION

Improved understanding of how demographic and WASH indicators vary within sub-national areas is increasingly recognized as important for targeting interventions and meeting the SDGs. Monitoring such indicators and measuring progress towards health and development goals requires reliable and detailed data.[139] However, spatially detailed information on demographic, socioeconomic, environmental, and health indicators in low-resource settings is usually only available at aggregated regional levels through nationally representative household surveys, such as the Demographic and Health Surveys (DHS), or national censuses. DHS data cover a wide range of indicators, including some useful for water and sanitation monitoring and impact evaluation. Data used to monitor progress towards national indicators, though, are inherently obfuscated to the first-level administrative unit, which may not provide an accurate picture of what is happening at sub-national and sub-regional levels.

Many organizations and researchers collect household-level data for their own purposes, but this data are not always made available to the public, and indicators are not always the same across projects, making it difficult to share information.[140] For example, rural sanitation data across countries and programs is not always consistent. WASHPaLS, a USAID-funded project, obtained datasets from six programs in Cambodia, Ghana, Liberia, and Zambia, covering a combined 40,482 communities before data cleaning.[141] Indicators from these datasets varied and not all communities included in these datasets were geo-referenced. The variation in indicators across WASH datasets makes it difficult to leverage household data for monitoring, evaluation, and research, and ultimately more than half of communities were removed from further analyses. Of 18 explanatory variables, only one was universally available and comparable in all four countries.

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.[142,143] The need for public access to high-resolution, geo-referenced data on contextual factors is increasing, while field data collection remains costly, time intensive, and potentially redundant.

To make household data more actionable, high-resolution maps can be generated from a smaller sample of point household data using GIS algorithms that perform interpolation. This statistical approach uses the existing data to predict values for nearby locations that lack data.[144] It can help to fill gaps in datasets where some information is missing. Various interpolation methods include kriging, inverse distance weighting, and spline;[145] the optimal method depends on the nature of the variable, the timescale on which the variable is represented, and the modeling surface. These powerful spatial analyses are commonly used for elevation, temperature, precipitation, and soil mapping; however, due to their complexity, they are not typically used by WASH actors. This is due in part to sparse data in developing and rural areas, where maps of even common variables are hard to come by. A few examples of large-scale efforts to interpolate contextual data at the national level exist. Smaller actors could leverage their survey data to produce interpolated maps at smaller scales in combination with other implementers who work in the same program regions.

DECISION STATUS QUO

WASH organizations often rely on a few large-scale efforts, such as DHS, MICS, JMP, and UN-Water Global Analysis and Assessment of Sanitation and Drinking-Water (GLAAS), to collect data on regional and national trends. Country-specific examples include national censuses and additional development programs such as Open Development Cambodia and the District League Table in Ghana. While some coordination mechanisms combine results and lessons learned, these are notoriously difficult to convene at a granular level and not always effective in influencing at-large implementation.[146,147] More high-resolution data maps are becoming available to implementers with the expansion of data extraction from satellite imagery, but often knowledge of these data sources and how to use them effectively for decision-making is poorly understood. Previous research emphasized certain variables, such as poverty, population, and education, which are useful indicators across development sectors.

Another approach is to aggregate and harmonize data from multiple household surveys after it has been collected. Projects such as the Central American Rural Water and Sanitation Information System (SIASAR) and the Water Point Data Exchange (WPDx) compile data from multiple sources. SIASAR was developed by the governments of Nicaragua, Panama, and Honduras.[148] It serves as a central, open platform where governments and CSOs can upload and share WASH data. It encompasses the full cycle from data collection, validation, analysis, and use through its suite of web-based applications. WPDx was developed by the Global Water Challenge in partnership with businesses, CSOs, the World Bank, UNICEF, and World Vision.[149] The platform allows users across the globe to upload data on water points. Using a set of rules to help standardize data, WPDx aggregates information and allows users to explore the data through their website. WPDx has aggregated data on about 250,000 water points in 25 countries from about 30 water data sources.

These solutions have some significant challenges. The main barrier for ensuring reliable comparable data is the validity of the data uploaded to the platform. This requires clear rules and standards for data collection, which may require additional resources in the form of ICT applications or capacity building for enumerators or users who input data. Additionally, some organizations may have restrictions on sharing data.

Program managers at civil society organizations (CSOs) do not always know where to most effectively target sanitation efforts or where the greatest need for a certain intervention exists. In addition, without detailed and reliable data, program managers may not understand existing situations in a given area (e.g., high disease prevalence). If location-specific data are required for program implementation, CSOs will often perform their own data collection. For example, a CSO professional noted, “for any program implementation, initial community assessment is critical.”[150] **Without improved household-level data, decision-makers are unable to tailor implementation solutions and decision-making to the area’s specific context.**

DEMAND

One interviewee noted, “[it is challenging] to develop a plan [or proposal] that is sound enough to be effective in [a short] timeframe and with limited data. Knowledge on the ground makes it easier to develop robust plans, and tools can be useful so that everyone can have access to the same information.”[150]

Resolution of data variables is a common issue for statistical modeling. Variables that are aggregated across very broad geographies, and thus have low within-indicator variability, can cause a model to become unstable and affect the reliability of results. Thus, high-quality interpolated maps can strengthen statistical analyses and predictive modeling: “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 or the reality of a situation but cannot replace an approach to a community.”[151]

Stakeholders also identified that often household data are not current enough to provide needed information. “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. An example is the Nakuru County investment plan in Kenya; the data in the final plan does not correlate with the current situation. There should be a system for continuous consultation and update of data for evidence-based decision-making.”[57]

OTHER DATA APPLICATIONS

Outputs from software application(s) developed for this use case could be applied to additional decision-making arenas. Research investments are not always leveraged to their full extent, and dissemination is a challenge for researchers. The ability to predict variable values where there are gaps could improve research cost-efficiency in the long run.

Maps of contextual factors are not limited to use within the sanitation sector. They can likewise be developed by those with other specialties or research foci. For example, WASHPaLS used a road accessibility map developed by an organization focused on malaria research. Interpolated maps of WASH household data can be used for other research/implementation needs, such as studies on nutrition, vaccinations, child stunting, literacy, etc.

EXISTING AND UPCOMING INNOVATIONS

Innovation 1: DHS Modeled Map Surfaces

Relevance: An example of an existing modeled map surfaces methodology that can be applied to achieve the use case objective.

The DHS Program provides a standard set of spatially modeled map surfaces from recent population-based surveys.[144] Each modeled surface is produced using standardized geostatistical methods, publicly available DHS data, and a standardized set of covariates across countries. Each map package contains a mean estimate surface, an uncertainty surface, and corresponding information on the model creation process and validation. The surface maps demonstrate the potential of geostatistical mapping techniques to produce interpolated surfaces from GPS-located DHS survey variables.[144]

The DHS modeled surface maps appear to be the only source which produces maps for multiple variables and across multiple countries. These maps, however, are not developed for every variable nor for every country, and it is unclear if updated maps will be produced with new surveys. These maps are possible because they are produced internally with access to non-obfuscated data. Other innovations are associated with peer-reviewed studies and maps, provided either as the finding of the paper (cholera, water scarcity), or as a proof of concept for an interpolation methodology. Interpolation (of various methods) can be done with GIS software (including ArcGIS and qGIS), although the user must determine how to format their data and select the optimal interpolation method. We did not identify any software application aimed at helping governments or smaller CSOs and implementers to understand how to interpolate their own data.

The following are examples of organizations using interpolation methods with DHS data:

- Researchers were able to observe disparities in latrine coverage of improved sanitation among different wealth categories in Kenya using data from the Demographic Health Survey (2007–2008). A series of spatial analysis methods including excess risk, local spatial autocorrelation, and spatial interpolation were applied to observe disparities in coverage.[152]
- Researchers created high-resolution poverty maps for low- and middle- income countries, including Kenya, Malawi, Nigeria, and Tanzania, using spatial interpolation and Demographic and Health Survey (DHS) Program data from 2017.[153]

Innovation 2: Demonstration of interpolation methods to create health- and wealth-related maps

Relevance: Existing methodology examples can be applied to achieve the use case objective.

One study combined environmental data and mobile phone data records for improved poverty prediction and mapping in Senegal.[154] Another study used the kriging method for epidemiological mapping of an influenza-like epidemic in France.[155] A third study proposed a method for interpolation of yearly local-level covariates of health status suitable for panel data analysis (a statistical method comparing cross sections of individuals over time) of the effect of health services.[156]

PARTICIPANTS

National governments (e.g., ministries of health): National government representatives would likely be at a management level at a ministry of health. They understand basic data summaries and statistics, and they are comfortable interpreting maps. Their time is limited, so they need a dashboard that is easy to interpret. National government representatives would use a proposed application in the following ways:

- Develop policies based on observable WASH outcomes.
- Prioritize funding and program implementation.

Implementing organizations (CSOs): Implementing agency representatives would likely be at a management level. They are familiar with the local context, although a data dashboard could enhance the depth of understanding. They are comfortable with interpreting data and maps and would have sufficient time to use applications. Implementing organizations might use applications in the following ways:

- Plan future projects based on levels of coverage of variables of interest.
- Design upcoming interventions based on variables of interest.
- Evaluate outcomes compared to other similar projects.
- Upload data from prior implementation efforts.

Universities and research organizations: Lead researchers, faculty, and upper-level students from universities and research organizations would engage with applications. They are generally familiar with managing and interpreting data and will have the technical understanding necessary to use applications independently. They have the time to use data throughout the research process. Researchers could use applications in the following ways:

- Design upcoming studies and target specific areas based on existing knowledge.
- Upload data from prior research, which can be used to provide guidance on implementation.

Developer: A developer's primary expertise is in developing models, conducting analyses, and providing data interpretations. Their time would be devoted to the development of applications during the building phase. Since data analysis skills across participants vary substantially, any applications would need to interpolate data with a high level of automation, and with little room for user interpretation or error. In addition, system maintenance will need to consider how data are collected and by whom to understand potential errors and to ensure data validity. A data scientist would need to be involved on an ongoing basis to maintain improve applications.

OUTPUTS

- **Web-based application** where users can enter data and produce maps of high-resolution WASH indicators of interest, including interpolated data

    - Map outputs will be cropped to the areas for which the point estimate prediction has an uncertainty below a default or user-selected threshold. Along with the map of indicator point estimates, the application will output a map of uncertainty so that future users of the map can determine how reliable the data are in their area.

- **Guidance documents** for use, survey design, and data management
    - Guidance documents will be geared toward program implementers with examples using familiar sanitation indicators. Survey design guidance will help implementers understand how to produce a reliable map for the indicator of choice and region of interest. For example, rare outcomes require more sampling (that is, the representation of an outcome or value within the dataset will affect the representation in the map output). Guidance will also offer advice on the geospatial spread of household clusters for sampling/surveying.

WORKFLOW

Step 1: Application development

The first step is to develop the standalone online application that users (e.g., sanitation implementers) could use to produce maps. It is important for the developers to emphasize user-friendliness and a step-by-step approach which begins with data formatting and determining interpolation approaches and suitability. The application will essentially perform the mechanisms in a mapping software (e.g., ArcGIS, QGIS), but via an online application to eliminate the need for software download and extensive training.

The team will also need to develop guidance for application use, survey design, and data management. Survey design guidance can promote methods to improve statistical power and support interpolation.

Step 2: Map development: Upload dataset

Users will upload a dataset with the available point data. Interpolating data into modeled surface maps removes any household identifiers. When aggregating data, however, steps should be taken to anonymize the data and remove any individual or household identifiers.

Step 3: Define model inputs

The user will need to define their model inputs and match their data to predefined categories. They will be able to select explanatory covariates from a list of preloaded global datasets. Geospatial covariates can partially explain variation in geospatial modeling, and their inclusion in the model allows for more accurate predictions across the map.

Step 3: Run data statistics (optional)

Three basic exploratory analyses can be built in to explore the characteristics of the raw data. The user can elect to download these plots for their records.

- Cluster-level observations: to show the location and observed values of the indicator for each geo-located survey cluster.
- Histograms: to assess the statistical distribution.
- Variograms: to summarize the spatial autocorrelation structure.

Step 4: Fit model

Fitting the model will happen on the back end. The model will parameterize to determine the set of values that leads to the best possible fit with the data.

Step 5: Validate model

Model validation will also happen on the back end. The model would use standard validation statistics to assess the predictive performance of the geostatistical model. The application will alert the user if model performance is too low.

Step 6: Download model outputs

The model will produce two map surface outputs, which the user can readily download.

- Point estimate surface map: plots the modeled point estimate value for each pixel.
- Model uncertainty surface map: summarizes the level of uncertainty associated with the values shown in the point estimate map by displaying the width of the 95% credible interval for each pixel value.[144]

These map surface outputs include estimates of the variable of interest at each location on a gridded surface (described in step 6 of the workflow). The modeled surfaces can be aggregated up to different administrative levels or other geographic areas to operationalize the maps for use in decision-making.

Step 7: Link map outputs to the Humanitarian Data Exchange

We encourage users to upload their maps (of values and uncertainty) to the Humanitarian Data Exchange. The Humanitarian Data Exchange is an open platform for sharing data across crises and organizations and allows the maps to be made open source for other researchers and implementers.

DATA SOURCES

Existing data:

- This application would generate interpolations from user-inputted data. Therefore, existing data are data that the participant has already collected or sourced (e.g., from DHS).

Data gaps:

- A training data set will be needed to develop the application.

BARRIERS

Uncertainty: The interpolation method brings a certain amount of prediction uncertainty. Uncertainty can vary across a modeled surface for several reasons, such as the sparseness of point location data, rareness of the estimated indicator, and the extent to which the model explains the variance. However, to address this, the uncertainty surface map helps users understand the robustness of an estimate at any given location on the map. When using the modeled surfaces for decision-making, it is important to consider the uncertainty of the estimate.

Data resolution: A key limitation is the mapping of urban areas. For national-level spatially modeled surfaces, urban areas tend to be predicted with relatively uniform values, in part due to the size of the final pixel resolution. This is because large urban areas typically exhibit heterogeneity for indicators at small scale, whereas rural areas may still demonstrate heterogeneity among predictive values at larger pixel resolutions. Specific conclusions related to urban areas at larger pixel resolutions should be considered carefully.

Data standardization: When preparing georeferenced indicator data for modeled surface maps, it is important to define and implement a standardized grid format. Geospatial data from different sources may come in a variety of spatial resolutions and geographic extents. To overcome this, it is important to standardize the data.

7. COMMUNITIES: UNDERSTANDING LOCAL CONTEXTS THROUGH COMMUNITY CLASSIFICATION

OBJECTIVE

To provide governments, donors, and implementing partners with granular, geo-referenced data on community classification (a composite metric of distance to roads and towns and travel time to large cities) to tailor rural sanitation or other interventions.

Decision: Stakeholders would use community classifications to determine which rural sanitation intervention, or mix of interventions, should be implemented in a given community depending on the likelihood of success and the costs of implementing the intervention(s).

DESCRIPTION

In a global context, improving effectiveness of rural sanitation interventions is critical to meeting SDG 6.2 targets (universal, adequate, and equitable sanitation access) and improving public health. Community-Led Total Sanitation (CLTS) is the most widely used rural sanitation intervention in developing settings. While investments in sanitation must accelerate to remain on track with SDG 6.2, increasing the cost-effectiveness of sanitation policies and programs is equally central to achieving this ambitious goal.

As governments, donors, and implementing partners collaborate to achieve district- and country-wide open-defecation-free (ODF) targets, WASH professionals increasingly acknowledge that different contexts call for different approaches to encourage the construction and sustained use of improved latrines. In their CLTS handbook, Kar and Chambers recommend that implementing organizations assess communities for their “challenge level” (based on community characteristics) prior to initiating CLTS interventions.[157] For example, the handbook considered communities more favorable for CLTS if they were small and remote, with wet conditions, high incidence of diarrhea, and no previous sanitation subsidy programs, among other factors.[157] The “Rethinking Rural Sanitation” Guidance (RRSG), developed by UNICEF, Plan International, and WaterAid, calls for situational analyses at both national and district/province levels to determine community typologies and guide the design of rural sanitation programs.[143] Still, the RRSG does not provide detailed methodologies for obtaining, integrating, or analyzing the necessary data to determine community typologies and the corresponding ideal programmatic mix. Previous research on contextual factors and CLTS program outcomes demonstrated the possibility of using existing datasets to identify local conditions that influence CLTS program outcomes.[141] Analyzing and applying information on local conditions can provide a basis for designing more cost-effective rural sanitation interventions.

DECISION STATUS QUO

There is little evidence of existing data analysis applications specifically for rural sanitation program planning and design. **Without granular, geo-referenced data on community classification, governments and CSOs will continue to select communities and implement rural sanitation programs without understanding the likelihood of achieving ODF status within a community and tailoring their programmatic approach. In addition, without a rapid community assessment prior to program implementation, it is difficult for implementers to accurately calculate projected costs.**

While the population in sub-Saharan Africa with basic sanitation services doubled from 2000–2017, estimates state that achieving universal coverage by 2030 will require the 2017 annual rate of increase to double.[1] A central challenge to meeting the SDG target is selection of the most appropriate sanitation option in each local context. These challenges are particularly stark in rural areas, where seven out of ten people lacked basic sanitation services as of 2017.

DEMAND

Aquaya received positive feedback from multiple stakeholders, including UNICEF, Water Supply and Sanitation Collaborative Council (WSSCC), Global Sanitation Fund (GSF), WaterAid, Global Communities, Organisation for Economic Co-operation and Development (OECD) International Development Statistics (IDS), World Bank, and Plan International, regarding the usefulness of an in-progress community classification app. Stakeholders felt the classifications could be very helpful for identifying which communities are hard to reach and therefore would have more costly sanitation interventions. Classifications could be used by governments and development partners when designing programs with donors. WaterAid, which was central to the RRSG, has also developed a community-level application to assign communities to one of the same typologies. A drawback to their application is the inability to apply the typologies at scale, which would “really help to solve challenges with regards to assessing large geographic areas.”[56]

Implementers and funders would benefit from greater coordination and an understanding of where and when different sanitation programs have already been implemented. This historical context is important for planning future programs in a specific geographic area: “It is important to document which programs are implemented where before starting any program.”[57] A program manager noted, “[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.”[58]

Other interviews gave additional feedback about Aquaya’s draft approach:

- “Overall [this application] is absolutely useful. Much more useful if you can apply at local levels and get a certain level of granularity for sub-national level.”[59]
- “I think that this [application] would be extremely useful. I love this kind of stuff. […] This is much better and easier to understand.”[60]
- “It’s a circle that we see over and over happening because community assessment is not done properly from the beginning. So that’s why I’m very supportive of this work. I always use these words, ‘wishful thinking’, but this is the right way to go, this is what we need, and what every country needs to do. This is also the way to empower and get the accountability from the government as well.”[158]

OTHER DATA APPLICATIONS

The community typology tool for sanitation intervention planning could be leveraged to serve other related purposes in high demand from WASH decision makers.

Application 1: Improve knowledge of prior experiences with past interventions, past research, and related outcomes (to interventions and research)

Decision: Governments (local or national), implementing organizations (CSOs), and funders could use granular, geo-referenced data on local conditions, in conjunction with past interventions, research, and outcomes, to a) prioritize future sanitation interventions, b) prioritize future sanitation investments, and c) track impact of investments.

Demand:

- "Granular (sub-district level) information on WASH access levels and history of sanitation interventions would be helpful... [It] would help prioritize areas for program implementation."[150]
- "[It is] difficult to plan a program when you do not know what has been done there before (e.g., what has succeeded or what has failed)."[159]
- "It would be very helpful to World Bank if there is a way to map WASH interventions, because this would guide [us] on which programs to implement/fund."[163]
- "Understanding current coverage and how it evolves, will always be important, so WSSCC knows where to focus the efforts, which programs are working, which are not working."[159]
- "Due to different stakeholder approaches on CLTS implementation and lack of data on what has been done, it becomes difficult to trigger communities' behavior. For example, when Helvetas Swiss Intercooperation wanted to roll out CLTS program, there was lack of data on previous CLTS interventions and their impact. [We] did not have any information on how the CLTS program would work, what to do, and what not."[164]

Application 2: Identify communities most vulnerable to climate change

Decision: Governments (local or national), implementing organizations (CSOs), and funders could use granular, geo-referenced data on local conditions, in conjunction with climate modeling, to identify a community typology for vulnerability to the effects of climate change.

Demand:

- "Climate change data are rarely analyzed or included in humanitarian frameworks and more talk is on WASH rather than environmental health. This is an area that can bridge development and humanitarian world[s] when designing strategies for a response."[165]
- "Data [can be used] to compare and define needs or areas of intervention that needs to be prioritized (e.g., vulnerabilities) [for] climate change."[165]
- "Granular data on rain patterns would improve servicing operations. For example, granular data can help to anticipate the need for more frequent cartridge emptying during a timeframe when it is expected to rain heavily. Modeling water movement throughout an informal settlement, for example, can also play an important role in deciding where to construct toilets."[66]
- "Climate change is a major challenge as one time there are floods and the next time there is drought.

[…] Data on the most appropriate water supply system for different geographic locations would therefore be very important. This would ensure that finances are channeled to the correct projects that would best benefit communities."[37]

- "There are possibilities for data for groundwater quality, depletion monitoring, and the impacts of climate change."[6]
- "Climate change [is] an interesting area that impacts water sources, storage and distribution network[s], and the household living conditions. However, data on climate change is not linked up with basic service provision during decision-making."[166]
- "There is an immediate need for better flood forecasting."[167]

Application 3: Understand progress towards achieving the SDGs

Decision: Governments (local or local or national), implementing organizations (CSOs), and funders could use district- or county-level, geo-referenced data on local conditions, in conjunction with information on progress towards the SDGs, to a) prioritize future investments based on SDG achievement, and b) coordinate development partners.

Demand:

- "Water sector investments should be guided by information and data-based strategies. That has not happened. One of the challenges is inadequate data. […] There is no information on the levels of SDG achievement across the country and progress over time."[38]
- "There should be a unified, comprehensive geo-database that feeds into JMP [WHO/UNICEF Joint Monitoring Programme]. It should be the same platform for all water and sanitation stakeholders."[38]

EXISTING AND UPCOMING INNOVATIONS

Innovation 1: WaterAid community-level survey tool (via mWater)

Relevance: This existing community classification tool for WASH programs could be applied to achieve the use case objective.

WaterAid has developed a community classification tool similar to Aquaya's (using the same typologies); however, it require surveying a community representative for use at the community level.[93] The WaterAid tool is being piloted in Niger and Rwanda. Further, the factors involved in the typology determination differ from Aquaya's proposed app, which fills a need for applying typologies over large geographic areas.

PARTICIPANTS

National governments (e.g., ministries of water and sanitation): National government representatives would likely be WASH officers within a ministry of water and sanitation. They are familiar with data summaries and statistics and comfortable interpreting maps. They have limited time to devote to reviewing data. National government representatives would use the community classification application in the following ways:

- Prioritize regions for rural sanitation programming and funding.
- Derive community data to inform national sanitation planning, guidance, and policies.

Local governments (e.g., public health offices): Local government representatives would likely be at a management level at a public health office. They are familiar with the local context and geography and interpreting maps, data summaries, and statistics. They have some time to interpret and understand the data to make decisions locally. Local government representatives would use applications in the following ways:

- Prioritize sub-regional areas for rural sanitation programming and funding. Ideally, ownership of applications resides with local governments in charge of sanitation in their respective jurisdictions.

Implementing organizations (CSOs): A typical representative is likely a regional manager for a CSO. They are familiar with the local context and geography and interpreting maps, data summaries, and statistics. They have some time to interpret and understand the data to make local implementation decisions. Implementing organizations would use the community classification application in the following ways:

- Select areas for rural sanitation programming and funding. More specifically, applications can be used to conduct community assessments and assess sanitation preferences or needs before program implementation (i.e., during the program proposal development stage).

Funders: A representative decision-maker is likely a grants manager for a large foundation or a contract project manager for a large multilateral funder. Funders are familiar with data summaries, statistics, and interpreting maps, but need a simple method to interpret the outputs of applications. They will have sufficient time to use applications to allocate funds to geographic areas in need. Funders would use applications in the following ways:

- Select rural sanitation programs to fund from those predicted to have the greatest likelihood of success.

Data scientist(s): Data scientists are data analysis experts familiar with sanitation in low- and middle-income countries. They gather data from the public sources, governments, or CSOs to develop applications. Their time would be devoted to the development of applications during the building phase, followed by opportunities to receive input on usefulness and functionality. They would also support maintenance of applications, as needed.

OUTPUTS

An online platform with granular, geo-referenced data on community classification. One example is the web platform that Aquaya has developed as part of the USAID WASHPaLS program.[168]

WORKFLOW

Step 1: Desk review

A data scientist would conduct a rapid desk review of existing guidance and toolkits on rural sanitation programming and design and data management considerations. The purpose of the desk review is to identify gaps where a community classification application can add value and leverage previous research on data use for sanitation programming. In addition, stakeholder interviews can assess demand for the community classification application and receive feedback on use and functionality aspects.

Step 2: Identify the primary variable

The first step in developing a software application is for the data scientist(s) to define variable value thresholds to facilitate pixel by pixel classifications using three primary variables: i) distance to roads, ii) travel times to cities, and iii) distance to towns, starting with a few locations.

Step 3: Create online interactive platform

Using the classification criteria, the data scientist will create maps for a few locations and load them onto an online interactive platform (map visualization only). Opportunities for improvement include gathering feedback from users (e.g., partners that provided data) on class definitions, usefulness, and interactivity or testing classification results with known communities.

Step 4: Expansion of application to select countries

The data scientist would next expand the online platform to include additional countries. This iteration of the online platform could include additional variables and user functionalities, such as the option to upload GPS coordinates and download national, sub-national, and community classification statistics. The platform will also include relevant plots and links to the RRSG-recommended implementation strategies tied to the community classification. At this stage, the data scientist(s) can again gather feedback from key users on the application usefulness and functionality.

Step 5: User testing

The team will identify a partner(s) to beta test the application within an actual program planning process. They will need to develop a testing protocol and interview guidelines. In addition, they will need to document uptake rates, use case applicability, and potential challenges for use at-scale.

Step 6: Expansion of application to include additional variables and user functionalities

The online platform may be expanded to supplement community classification with secondary datasets on literacy, vegetation coverage, access to drinking water, and soil types, where available. Once a polished application exists using base classifications (established only with distance to roads, travel time to cities, and distance to towns), the online platform will be expanded to include increased functionality and variables, with changes first implemented in the initial countries. If a new weighted matrix is developed, it is important to gather feedback from key users on the incorporation of the new variables. Additional data layers available for large geographic areas, such as land cover and soil type, can be added, allowing users to filter and toggle layers of the interactive maps on and off. Additional functionalities to be considered given input from stakeholder interviews include adjustment of the decision matrix and re-classification of community typologies on-the-fly. During application expansion, typologies could optionally be refined with additional variables embedded in the classification process. The data scientist should also revisit the long-term hosting strategy and finalize any outstanding decisions that influence application development.

Step 7: Application of additional updates to the full platform

Once the final functionalities and variables have been decided, updates should be rolled out to the remaining countries, matching the scope of Step 6. At this stage, the data scientist would again gather user feedback from key relevant stakeholders to finalize the software application. Finally, the application would be shared publicly.

Step 8: Dissemination and promotion

The development team should lastly disseminate and promote use of the application in the sanitation sector. A dissemination strategy should be developed by the data scientist, in collaboration with key users. Dissemination events might include sector workshops and conferences.

DATA SOURCES

Existing data:

- The USAID WASHPaLS research indicates that it is possible to use existing datasets to identify local conditions that influence CLTS program outcomes.[141] Analyzing and applying information on local conditions provides a basis for designing more cost-effective rural sanitation interventions. In response to WASHPaLS research findings, partners have requested guidance on how they might use the results for future planning of rural sanitation programs. Approximately 13 publicly available datasets, including a newly developed rural typology, will be leveraged for use on a single, user-friendly sanitation planning application called SanPlan.[168] Covering more than 10 countries, SanPlan will offer five analysis features to help users explore highly localized (at least 5-km) spatial data.
- The primary data for the application includes country-wide variables on a community's distance to roads, travel time to cities, and distance to towns.
    - Distance to roads can be sourced from Open Street Maps.[169] This raster layer is created in ArcGIS/QGIS using road shapefiles extracted from Open Street Maps and an administrative boundary shapefile. The layer calculates the straight-line distance between the centroid of each pixel and the nearest road feature and assigns that distance to the pixel. Prior to creation of the distance layer, the Open Street Maps shapefile is altered to only include major roadways (trunk, primary, secondary and tertiary).
    - Travel time to cities is an existing global raster file created by the Malaria Atlas Project.[170] No data cleaning is required.
    - Distance to towns can be sourced from a population density dataset (continental raster file) created by WorldPop.[100] Hotspots of dense populations can be used to create a layer of towns which are likely to have viable sanitation markets. The methodology outlined above for "distance to roads" can also be applied to creating a "distance to towns" map layer.
- Secondary datasets for the application may include: community size, latrine coverage, ODF status, GPS coordinates, literacy, vegetation coverage, access to drinking water, and soil type.
- Administrative boundary shapefiles can be sourced from Diva-GIS.org.[171]
- Existing options that offer interactive global mapping include AtlasAI, a technology startup that develops data products to support economics, agriculture, and infrastructure improvements in low- and middle-income countries.[172] Data layers are created using remote sensing and artificial intelligence, for example to generate detailed insights on poverty, crop yield, and economic trends across Sub-Saharan Africa. The map layers are high resolution (2-km pixel) and offered at the continental scale. Products are offered as data downloads or as an interactive web map.
- Sustainable Development Solutions Network has developed an interactive website providing visual representation of all UN countries' performance by SDGs.[173] The dashboards help to identify countries that require particular attention for early action. 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.[174] Interactive features include the ability for users to build their own plots, download data, and export plot images.

- The Socio-Economic Atlas of Kenya provides a visual illustration of the geographic and socio-economic data pulled from the national census.[175] The application allows users to change granularity of the map to see effects at varying administrative scales.

Data gaps:

While many contextual factors have been shown to impact the suitability of one sanitation intervention over another, datasets must be available at high resolution and at a large scale to be incorporated into this project. Point data, such as household or community survey data, cannot be included unless data are interpolated and meet a threshold for certainty. The following factors were identified in the RRSG, the CLTS Handbook, and a USAID desk review, as important sanitation considerations for which suitable datasets are not yet available:[142,143,157]

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
- Prevalence of rented accommodation

BARRIERS

Assumptions: The largest challenge in developing classification maps that span many countries is the assumptions embedded within the classification "rules," particularly as the rules are applied in countries where understanding of sanitation issues may be lacking. This risk could be mitigated by beta testing the maps with as many implementing partners as possible, to determine where approximations fit and where rules need to be modified. The final application would allow users to modify rules, thereby employing their expert knowledge of CLTS performance and local context to customize the classifications.

Data gaps: Potential challenges include data availability (i.e., no data exist for given variable of interest), accessibility (i.e., data exists but is inaccessible for political or administrative reasons), granularity (i.e., data not available at the desired administrative level), or inconsistency (i.e., data not available for all administrative units). The desk review prior to application development can help to identify which variables might be subject to these challenges.

Buy-in: User buy-in and operation of the application may prove challenging. To mitigate this risk, a guidance document should be developed to accompany the application. Training events may also provide an opportunity to increase buy-in.

8. COMMUNITIES: TARGETING THE POOR AND VULNERABLE

OBJECTIVE

To provide governments (local or national), service providers (e.g., utilities), CSOs, and funders with high-resolution information on poverty levels and creditworthiness to target WASH subsidies and loans.

Decision:

- Stakeholders can use information on poverty levels to determine which households should be eligible for pro-poor financial support such as piped water connections, subsidized water tariffs, and sanitation products and services.
- Stakeholders can use information on creditworthiness to determine which households are sufficiently creditworthy to take a loan for purchasing a toilet or a piped water connection.

DESCRIPTION

Though access to safe water and sanitation 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 from these services.[176] The imperative set by the SDGs to "leave no one behind" requires dedicated strategies to improve water and sanitation access among the poorest two quintiles (40%) of the population. Subsidies and loans are promising financial instruments, if they can be targeted at the households that most need them. Loans are appropriate for households that can afford WASH infrastructure upgrades 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). Subsidies are common in the WASH sector, but they are poorly targeted and thus largely ineffective: they primarily benefit high-income groups and often fail to reach the very poor.[177,178] This is largely due to the absence of effective methods for service providers or CSOs to identify the poorest households. WASH loans are also increasingly prevalent, but it is difficult for lending institutions to correctly identify households who can truly afford a loan and are likely to repay.[179-181]

DECISION STATUS QUO

Subsidies – Governments and CSOs typically use three types of approaches to target subsidies:

- National poverty identification systems: A few countries have a nationwide system to identify households meeting specific poverty criteria (e.g., IDPoor in Cambodia, Below-Poverty-Line cards in India, Livelihood Empowerment against Poverty [LEAP] in Ghana). Programs led by the CSO, iDE, in Cambodia and the government-led Total Sanitation Campaign in India have relied on these systems to target toilet subsidies.[182] However, because these systems rely on poverty criteria defined at the national level (i.e., combining rural and urban areas), they may not always provide sufficient sensitivity at the local level.[183] For example, if a region or city is wealthier than the country average, the national identification system may only recognize a very small proportion of households as "poor" and may not be useful for a utility aiming to target subsidies. Conversely, if an area is substantially poorer than the national average, the proportion of households categorized as "poor" may be too large and impractical for use by CSOs and local governments. In some cases, these systems are not yet consistently available throughout the country; for example, in Ghana, the LEAP program has not yet been rolled out to every community. Even limited national poverty identification systems are an exception rather than the rule, as most countries do not have a centralized, standardized process to track poor households.

- Participatory approaches: Participatory approaches rely on the input of community members to identify the poorest households. A number of rural WASH programs have used this approach to target subsidies.[78,184] Participatory methods generally have a high accuracy, but they are resource intensive and difficult to scale up quickly.[185]
- Increasing block tariffs: Water utilities typically apply block tariffs, wherein customers only pay a higher volumetric price if their consumption exceeds a designated use threshold. The rationale behind consumption-based subsidies is that poor consumers should have lower water usage, which will result in lower water tariffs. However, block tariffs have been ineffective in low-income countries because: i) they only apply in areas served by the utility, which usually exclude the poorest areas, and ii) wealth doesn't correlate well with water usage, as poor households often have larger household sizes and/or share a water connection with neighbors.[186] As a result, subsidies delivered through block tariffs do not perform better than randomly targeted subsidies.[187]

Cost-effective, scalable approaches for governments and CSOs to identify poor households at the local level would allow decision-makers to ensure that WASH programs reach the most vulnerable population segments.

Loans

Microfinance institutions (MFIs) rely on credit scores to allocate WASH loans. A credit score is a number representing the likelihood of a potential borrower to repay a loan. If the score is too low, the MFI will not give out a loan. To determine a person's credit score, MFIs typically rely on software that processes information on prior banking history, income and expenditures, assets, or social media data. Such software is often expensive (for example, a LenddoScore license costs 70,000 USD per year) and may not be appropriate for evaluating borrowers in low-income countries who have limited-to-no prior banking history or social media activity. **Without high-resolution data on creditworthiness, the most poor and vulnerable households will not be able to access credit to make necessary improvements in water and sanitation access.**

DEMAND

Through conversations with stakeholders, we identified demand for improved methods of identifying poor and vulnerable households for effective subsidies. A professor at the Malawi Polytechnic University indicated that a systematic, rapid software application to identify the poor would help allocate sanitation subsidies in urban areas more effectively. In the absence of robust data, subsidy programs rely on traditional chiefs to decide which households qualify, an approach that lends itself to bias and corruption.[56] Program managers in several international organizations (e.g., UNICEF, WaterAid) have called for data-driven approaches to identify the poor.[158,188] One interviewee reported that municipal governments would like to target subsidies at the poorest quintile, but identifying these households at the local level requires intense data collection. A World Bank economist[189] noted, "protocols for targeting subsidies are needed everywhere," adding that the challenge is not only to collect poverty data, but to collect it as cost-effectively as possible."

In 2018, the Government of Ghana issued guidelines promoting targeted subsidies in the sanitation sector. However, implementing organizations need more specific and operational guidance on the targeting process. A 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.”[37] The Water Sector Trust Fund (a Kenyan State Corporation that provides WASH grants and financial assistance) conveyed that identifying urban low-income areas (LIAs), their priority investment target, is difficult because LIAs are not static: income levels evolve over time, and rapid urbanization leads to LIA expansion or creation.

In addition, we identified demand for improved methods to target loans. The CSO iDE in northern Ghana offers loans to households that want to purchase the SamaSama pour-flush toilet. Program managers reported that existing credit-scoring software does not accurately predict repayment behaviors among their customers.[190]

OTHER DATA APPLICATIONS

Application 1: Develop a database of the poorest households

- Decision: Governments and implementing organizations could use a local-level directory of the poorest 20–40% of households, or a cost-effective, rapid approach to a) target other social benefits (health insurance, lower school fees, electricity, fertilizers), and b) allocate aid in emergencies.
- Demand: As previously described, demand exists for a simplified method to identify poor households. Having an easily accessible and understandable directory would enhance the ability of governments and implementing organizations to quickly direct social benefits and aid to the poorest households.

EXISTING AND UPCOMING INNOVATIONS

Innovation 1: Predictions of poverty status using simple household characteristics.

Relevance: “Smart” surveys represent an existing technology that could help to achieve the use case objective.

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.[191,192] World Bank researchers have demonstrated this approach in Malawi and Indonesia and are incorporating AI into concise, dynamic poverty identification surveys.[193] The USAID IUWASH PLUS program in Indonesia showed that a short, rapid survey provided sufficient information to map poverty clusters at the city level.[194]

Innovation 2: Poverty mapping using remote sensing

Relevance: High-resolution poverty maps represent an existing technology that could help to achieve the use case objective.

Several studies have shown that satellite imagery (particularly night-time images) and mobile phone data can be used to map poverty at the sub-national level.[195–197] Researchers have also applied satellite imagery to identify poverty levels at very high resolution (household level or 100 m x 100 m), although the performance of these models could be improved.[198]

Innovation 3: PulseSatellite

Relevance: Collaborative satellite image analysis application that relies on AI-based analytical methods. It is an existing technology that can be applied to achieve the use case objective.

PulseSatellite can be used to map refugee settlements and identify and classify structures.[199] PulseSatellite can also be used for rapid flood mapping, which is essential for emergency response planning. High levels of accuracy are possible with their machine-learning methods.[199]

PARTICIPANTS

Data analysts: Data analysts dedicated to development activities are experts in data analysis and modeling and generally familiar with developing country contexts. During engagement, the data analyst will be devoted to product development. The data analyst would be responsible for developing and updating two types of products:

- **Short, "smart" surveys** that rely on a minimum number of questions to accurately predict a household's poverty status and creditworthiness. Data analysts rely on artificial intelligence, available nationally representative surveys, and data from micro-finance institutions to develop these "smart" surveys.
- **Poverty maps** that display heterogeneities in poverty incidence at the district or municipal level and rely on satellite or aerial imagery to remain up to date. Data analysts will need data from the "smart" surveys to develop these maps.

Local governments (e.g., district or municipal governments): Local government representatives are likely field staff or office-based staff. Given their role, their experience with data interpretation varies. The local government representatives are responsible for identifying poor and vulnerable households at the local level, and thus could directly contribute to the development of proposed applications and benefit from its output.

- Field staff administer a short, "smart", mobile-based survey to households (maximum 10 questions). Field staff are trained in community engagement and mobile-based survey methods. Although the survey would be subject to reporting bias, questions could be worded carefully to avoid this, and some observable information could be verified by field staff during the household visit. The survey would automatically generate the household's poverty status using an embedded predictive model.
    - In rural settings, it may be possible to survey all households. Depending on the survey outcome, field staff could give households a "poverty card."
    - In urban settings, it may be more realistic to survey a representative sample of households.
- Office-based government representatives can review the data in real time on an online dashboard, including the number of poor households and their geographic distribution. They can also download the list of poor households by locality, or the list of localities with the highest poverty density. Finally, they can view and update poverty maps. These staff are familiar with data summaries and statistics and are comfortable interpreting maps.

Implementing organizations (e.g., small water or sanitation enterprises, utilities, CSOs): Implementing agency representatives are likely at a management level within their organizations. They are familiar with data analysis and interpreting maps. They will have limited time to interpret data and make decisions based on software outputs. Implementing organizations would use proposed applications in the following ways:

- Use data provided by the local government to allocate subsidies. For example, a small water enterprise may give poor households a water debit card with free credits, or a CSO may give them a voucher for a subsidized toilet.

- In urban areas, a utility may decide to subsidize piped water connections in locations with a high poverty density.

Service providers – public or private (e.g., market-based sanitation organizations, utilities): Service provider representatives will likely be at a management or project implementation level. They are generally very familiar with interpreting data, but are not familiar with administering questionnaires. They will have limited time to interpret the data and make decisions. Service providers use applications in the following ways:

- Assess a household's creditworthiness. Customer service staff administer the short survey with prospective borrowers. The credit score determines the household's eligibility for loans for products such as piped connections and toilets.

National governments and funders (e.g., ministry of water and sanitation): The national government or funder representative is likely a project manager. They are familiar with data summaries, statistics, and interpreting maps, but need a simple method to interpret program effectiveness. They will have limited time to interpret the results. These parties could use applications in the following ways:

- Require service providers and grantees to report their performance with respect to service coverage amongst the poor.

OUTPUTS

- A mobile-based survey automatically generating poverty status or loan eligibility status.
- A web-based dashboard providing local-level poverty maps, lists of poor households or localities, and summary statistics on the number of poor households, with comparisons across areas. The dashboard may be integrated with other data management platforms that the local government or utility uses. The data will be accessible globally.

WORKFLOW

App development: smart survey to assign poverty status

Step 1: Data collection

The team would collect data from nationally representative survey (e.g., DHS,[200] MICS,[25] or World Bank's Living Standards Measurement Study [LSMS][201]) and apply machine learning to identify a small number of household characteristics that accurately predict poverty status.

Step 2: Survey creation

The team would create a short survey including questions on these household characteristics, and embed the predictive model so that the survey automatically generates a poverty status.

Step 3: Survey validation

The team would validate the smart survey against the outcomes of participatory approaches such as community consultation (in a small number of locations).

Software application development: local-level poverty maps

Step 1: Data collection

The team would collect data on recent census-based poverty estimates,[202] or if not available, administer the smart survey above.

Step 2: Map creation

Data scientists would create a map of poverty prevalence and use GIS to delineate areas that qualify as "poor." If the data collected in Step 1 is too sparse, this may require applying interpolation techniques.[203]

Step 3: Combine survey map with satellite imagery

Data scientists would combine the survey-based poverty map with satellite or aerial imagery. Using artificial intelligence, they would train an algorithm to recognize poor areas on satellite or aerial imagery. The team will need to ground-truth the algorithm in a few locations.

Step 4: Application of algorithm

The data scientist would then apply the algorithm to satellite or aerial imagery in other locations to map poor areas. They would have to re-process satellite or aerial imagery at least annually to update the initial poverty map.

Software application development: smart survey to assign credit score

Step 1: Collect customer data

The team would collect customer data from a financial institution offering WASH loans. Since customers with a high likelihood of repayment are over-represented in these data, it may be necessary to augment the dataset by offering loans to other types of customers for a short period.

Step 2: Identify customer characteristics

A data scientist would apply machine learning to identify a small number of customer characteristics that accurately predict loan repayment.

Step 3: Create customer survey

Using priority characteristics, the team would create a short customer survey and application that automatically generates loan eligibility status.

Step 4: Survey piloting

The team would pilot the survey for 1–2 years with an organization providing WASH loans to validate, test, and refine the application.

DATA SOURCES

Existing data:

- Nationally representative surveys, such as DHS,[200] MICS,[25] or surveys on income levels(census data may or may not be up-to-date due to the frequency of data collection).
- Government reports on how to derive poverty estimates from census data.

- Customer data from micro-finance institutions.
- Satellite and aerial imagery available on platforms such as OneAtlas[204] or Digital Globe.[205]
- In some countries (e.g., Kenya), maps delineating urban low-income areas are available; however, these are not frequently updated or maintained.

Data gaps:

- Nationally representative surveys on income levels or poverty proxies are unlikely to be available in every country.
- High-resolution satellite imagery (i.e., pixel size of 50 cm or less) is not uniformly available in low-income countries.
- Micro-finance institutions may not have sufficiently detailed customer information to derive a predictive model of loan repayment.

BARRIERS

As noted above under "data gaps," nationally representative surveys on income levels may not be available or may be outdated. In addition, high-resolution satellite imagery is not uniformly available in low-income countries. In some areas, satellite imagery may be of insufficient resolution or outdated. In such cases, it may be necessary to complement satellite imagery with aerial images or drone images. Lastly, customer data from microfinance institutions may be difficult to access as they are often proprietary or protected by privacy laws.

9. PROGRAMMING: EVALUATING IMPACTS

OBJECTIVE

To provide national governments, implementing organizations (CSOs), and funders with alternative methods to determine WASH program impacts.

Decisions:

- Stakeholders can use this method to understand if projects met their intended goals.
- Stakeholders can use this method to decide which programs should be continued, scaled up, replicated, or discontinued.

DESCRIPTION

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. Nevertheless, it is critical in resource-constrained settings for governments, implementing organizations, and funders to understand and compare the cost-effectiveness of different interventions. Program evaluations typically occur towards the end of the program or after itscompletion; however, implementers and decision-makers have indicated they would benefit from real-time indicators and predictive analytics.

DECISION STATUS QUO

To measure impact, researchers typically rely on a control group, which provides a "counterfactual" to the intervention, to understand about what might have happened if no intervention was applied. Randomized control trials (RCTs) are considered optimal for comparing control vs. treatment groups, but these studies are costly, time-consuming, and sometimes unethical or unrealistic in development settings.[206] An alternative to RCTs are observational studies, which use historical data on naturally occurring (unplanned) experiments or collect data on scenarios wherein similar cases receive different "treatments." Observational studies are often criticized for being more susceptible to bias, as researchers may select which cases to consider as part of a study retrospectively, or the intervention status may be related to other underlying differences.

Machine-learning techniques offer opportunities to reduce the costs and complexity of RCTs and reduce the risk of bias in observational studies by using existing datasets and algorithms to identify naturally occurring counterfactuals or control cases. In addition, applying big data sets for program evaluation allows users to 1) explore additional variables or outcomes that might not be included in an initial theory of change, 2) observe population-wide effects, or 3) measure longitudinal changes more efficiently.

DEMAND

Although large implementing organizations may have the funding and capacity to design proper monitoring and evaluation systems for their programs, how success is ultimately measured remains in question. One international CSO interviewee mentioned, "Most donors, even implementors, don't have a realistic idea about what reasonable success could even be for a program."[207] In addition, many smaller organizations struggle to evaluate their efforts. A manager focused on providing pro-poor access at an urban Kenyan water utility noted, "NAWASSCO has 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."[57] Rural WASH programs also need improved evaluation systems, as illustrated by an interviewee from Helvetas Nepal: "[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. This kind of data will support design and implementation of WASH interventions."[164]

Not only can post-implementation evaluation be challenging, but often it can be difficult to understand the status of implementation progress, as mentioned in an earlier use case: "One 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."[160]

Lastly, investors desire better evaluation data to identify and prioritize investments in effective programs. Unfortunately, current limitations of evaluation data and methods make it difficult to advocate for high-impact programs. Another interviewee 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."[38]

OTHER DATA APPLICATIONS

Machine-learning techniques could be useful for evaluating programs in progress. Once factors associated with success are identified, models can be used to analyze data in real time and predict whether or not a program is on track to reach its intended goals.

EXISTING AND UPCOMING INNOVATIONS

Innovation 1: Machine-learning techniques analyze existing, "passively" collected data to monitor and evaluate program results compared to a baseline

Relevance: An existing low-cost method for evaluating programs.

The United Nations Millennium Development Campaign in partnership with the WSSCC and UN Global Pulse used social media data to evaluate the effectiveness of a sanitation campaign.[208] It relied on natural language processing techniques to evaluate perceptions and sentiments about sanitation. Correlating spikes in social media posts with relevant sanitation campaigns such as World Toilet Day helped to quantify the impact of campaigns over time. By showing how the volume and content of public discourse around sanitation changed over time using social media analytics, the study provided a baseline that could be used to monitor the effectiveness and reach of a communications campaign in real time. The analysis also revealed increasing public engagement around gender and sanitation.

Innovation 2: Machine-learning techniques analyze longitudinal program data to identify naturally occurring experiments and interventions or treatments that have a desired impact

Relevance: An existing method of combining data science with traditional program evaluation methods, which could contribute to the use case objective.

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 in the system.[209] Program experts and data scientists worked together on a blended approach of more traditional program evaluation methods and data science. They developed an algorithm that matched cases in terms background and context, while varying the types of "treatment" received by each child to understand which led to the desired impact.

Innovation 3: Analysis of publicly available datasets to identify comparative cases, or counterfactuals

Relevance: This innovation presents an alternative method of using data science to evaluate programs.

Stanford University's Water, Health, and Development program is leading the monitoring, evaluation and learning strategy for the Conrad N. Hilton Foundation's Safe Water portfolio. The Safe Water projects aim to increase access to safe water services.[210,211] To evaluate performance of projects in 12 focus districts, Stanford and the Hilton Foundation sought to compare changes in the project districts to other, similar districts that did not receive support. To enhance rigor, data on the comparison districts was gathered to ensure they were truly similar to the project districts. Stanford sought existing national and global datasets on a number of factors hypothesized to influence access to safe water, such as population density, poverty, terrain, etc. The use of existing datasets to characterize a counterfactual scenario eliminated the need for extensive primary data collection in multiple districts where program activities were not taking place.

PARTICIPANTS

Implementing organizations (e.g., national governments, CSOs): Stakeholders at implementing organizations would likely be at a management level. Project managers within governments or CSOs likely have extensive experience working with program goals and interpreting data to assess outcomes. They understand reports and statistics, but likely need the support of a data scientist to execute analyses. Project managers might use the use case outputs in the following ways:

- Understand how successful their programs were and why. Factors identified as promoting success can potentially be applied to future implementation in the same or similar contexts.
- Identify which programs should continue and/or be scaled up, versus those that should be discontinued and not replicated. This process should improve program fund utilization.
- Identify other geographic or societal contexts in which similar programs may be successful.

Funders: Stakeholders at donor organizations decide where funds should be allocated. Managers typically have limited skills in WASH data management and interpretation outside of financial figures. They often have limited time to spend understanding data to inform funding priorities. They can understand reports and basic statistics and interpret maps. Funders could use data in the following ways:

- Prioritize programs for funding depending on the projected likelihood of success, as determined by contextual factors, program goals, and success outcomes of comparable projects.
- Validate success of previously funded work to justify additional revenue.

Data scientists: Data scientists are individuals (or teams) that develop algorithms and machine-learning techniques for use in program evaluation. Developers have advanced computer programming and design skills. They are familiar with machine learning, but might need guidance on understanding WASH programming and its goals and outcomes. Their time would be devoted to the development of the algorithm during the initial phase, followed by maintenance and upgrades.

OUTPUTS

Using data science approaches for program evaluation requires an iterative approach to produce a unique algorithm that can support identification of specific project goal achievement. Since this needs to be done in conjunction with program or subject matter experts, the initial phases of this approach should not be automated.

Once factors for program success or expected outcomes are identified through the evaluation process, these can be turned into goals and targets for future programs and applied within predictive analytics. Ongoing monitoring can be automated by integrating predictive analytics into an online dashboard that would flag to implementers or funders when projects are not meeting specific targets.

WORKFLOW

Step 1: Develop theory of change

The team should first develop a theory of change that hypothesizes outputs and outcomes of a project, and the indicators or variables that affect outcomes. It needs to factor in different types of implementation strategies and goals.

Step 2: Identify appropriate dataset(s)

The team should next identify existing datasets and determine if additional data needs to be collected. Data should include outcomes or proxies of interest (e.g., satellite imagery of roof materials, social media data on perceptions or sentiments), as well as demographics, location, contextual factors, etc.

Step 3: Develop training dataset

The team should then develop a training dataset from existing and newly collected data to identify patterns or combinations of factors leading to the expected outcome.

Step 4: Execute machine learning

To analyze counterfactuals, the team can identify matched experiments (i.e., two cases in the dataset that have similar background or contextual characteristics, but that received different treatments). Alternatively, they can compare performance of one case over time relative to a baseline, or the performance of that case before and after the program intervention.

Step 5: Evaluate outcomes

The team must last assess outcomes of matched groups and quantify the impact. Additional feedback from program experts may need to be incorporated into the machine-learning method in an iterative process to optimize results.

DATA SOURCES

Existing data:

- Machine-learning techniques could be applied to a variety of data inputs, including program-collected data, "traditional" datasets such as large-scale national surveys, or "big" data such as:
  - User-created data (e.g., on social media)
  - Transactional data (purchase information, application usage, other passive data)
  - GPS/location data and satellite imagery
- Unlike data used for statistical analysis, big data applications can more easily combine quantitative and qualitative data analysis. In addition, there are fewer requirements for data to be standard or shaped (i.e., it does not matter if data are skewed).

Data gaps:

- Use of existing data sources for program evaluation rather than primary data collection may downplay or delay collection of new information helpful for evaluation.
- Remote and hard-to-reach populations may still lack data, especially if they are sourced using technology that is not easily accessed by these populations.

BARRIERS

Bias: Big datasets that rely on technology for data collection may have inherent biases. Certain vulnerable groups, such as the ultra-poor or women, may not have equal access to technology like mobile phones or internet access.[209] The quality of datasets should always be evaluated before use to ensure it is representative or flag its limitations.

Informed consent: In typical experimental studies, participants are informed of the research goals and have the opportunity to opt-out of the study. Such informed consent for additional analysis of secondary datasets is not possible, but arguably poses significantly less risk to individual human subjects.

Correlation versus causation: Training machine-learning techniques without a theory of change can conflate correlation with causation. It is therefore important for program experts to work closely with data scientists to iterate on the theory of change throughout the analysis.

Collaboration: Program evaluators and data scientists often operate in silos with different skillsets, and it can be challenging to bring the two fields together for these types of projects.[209]

References

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

2. 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.

3. Damania, R., Desbureaux, S., Rodella, A.-S., Russ, J., & Zaveri, E. (2019). Quality Unknown: The Invisible Water Crisis.

4. 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

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

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

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

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

9. 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).

10. 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.

11. 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

12. 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

13. 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.

14. 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. Scientific Investigations Report 2012-5065.

15. 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

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

17. Kenya, K., Senozetnik, M., Klemen, K., & Mladenić, D. (2018). Groundwater Modeling with Machine Learning Techniques: Ljubljana polje Aquifer. In: Conference on "Insights on the Water-Energy-Food Nexus." https://doi.org/10.3390/proceedings2110697

18. 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

19. IGRAC. (2021). Global Groundwater Monitoring Network. https://ggis.un-igrac.org/view/ggmn

20. IGRAC. (n.d.). Groundwater Quality. https://www.un-igrac.org/areas-expertise/groundwater-quality

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

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

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

24. World Business Council for Sustainable Development. (2019). India Water Tool 3.0. https://www.indiawatertool.in/

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

26. UNICEF. (2020). MICS Methodological Work: Water Quality. https://mics.unicef.org/methodological_work/3/WATER-QUALITY

27. WHO/UNICEF Joint Monitoring Programme (JMP) for Water Supply and Sanitation. (2021). Water quality monitoring. https://washdata.org/monitoring/drinking-water/water-quality-monitoring

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

29. 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

30. 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.

31. Interview with Roland Liemberger, Independent Non-Revenue Water Consultant, formerly Miya. (2020).

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

33. 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

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

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

36. WASREB. (2016). Tariff Guidelines.

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

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

39. 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/

40. Schneider Electric Global. (2021). EcoStruxureTM Geo SCADA Expert. https://www.se.com/us/en/product-range/61264-ecostruxure-geo-scada-expert/#overview

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

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

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

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

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

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

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

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

49. Leakmited. (2020). LEAKMITED: Water infrastructure management using satellite images and AI. https://www.leakmited.com/

50. 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

51. 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/

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

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

54. FSM4. (2017). FSM4 Case Studies.

55. 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: an experimental comparison of latrine pit-emptying services. Journal of Water, Sanitation and Hygiene for Development, 10(4), 744–755. https://doi.org/10.2166/WASHDEV.2020.060

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

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

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

59. Furlong, C., & Jooust, L. (2016). SFD Report Kisumu, Kenya.

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

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

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

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

64. Interview with Rick Johnston, JMP/WHO Officer. (2020).

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

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

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

68. Pula. (2021). https://getpula.com/

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

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

71. 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

72. 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

73. 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

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

75. 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

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

77. 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

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

79. 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

80. 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

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

82. 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

83. 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

84. 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

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

86. McGill University. (n.d.). Chapter 6: Essential Factors for the Provision of Sanitation Systems in Coastal Communities. https://www.mcgill.ca/mchg/student/sanitation/chapter6

87. Melo, J. C. (2005). The Experience of Condominial Water and Sewerage Systems in Brazil: Case Studies from Brasilia, Salvador and Parauapebas.

88. Seyoum, S. (n.d.). Type of Sewer Systems.

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

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

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

92. IRC. (n.d.). WASHCost. https://www.ircwash.org/washcost

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

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

95. 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

96. 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

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

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

99. CACTU$ Costing. (2020). Costs of Citywide Sanitation. WASH Research Group, School of Civil Engineering, University of Leeds.

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

101. NASA Jet Propulsion Laboratory. (2021). Shuttle Radar Topography Mission.

102. ISRIC. (2020). Soil property maps of Africa at 250 m resolution. https://www.isric.org/projects/soil-property-maps-africa-250-m-resolution

103. British Geological Service. (2020). Download digital groundwater maps of Africa. https://www.bgs.ac.uk/research/groundwater/international/africanGroundwater/mapsDownload.html

104. 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

105. 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

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

107. 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

108. 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

109. 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

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

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

112. 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

113. 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

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

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

116. 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

117. 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

118. 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

119. 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

120. 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

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

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

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

124. Walter Reed Biosystematics Unit (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

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

126. 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

127. 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

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

129. 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

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

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

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

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

134. 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

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

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

137. 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

138. 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

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

140. 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/10.1596/978-1-4648-1197-5

141. 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

142. USAID. (2018). An Examination of CLTS's Contributions Toward Universal Sanitation.

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

144. 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.

145. Esri. (n.d.). Comparing interpolation methods. Retrieved December 2, 2021, from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/comparing-interpolation-methods.htm

146. 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.

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

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

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

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

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

152. 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

153. 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

154. 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

155. 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

156. 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.

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

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

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

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

161. Interview with Michael Gnilo, UNICEF. (2020).

162. Interview with Eduardo Perez, Global Communities. (2020).

163. 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).

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

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

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

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

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

169. OpenStreetMap. (2021). https://www.openstreetmap.org/

170. The Malaria Atlas Project (MAP). (2021). https://malariaatlas.org/

171. DIVA-GIS. (2021). https://www.diva-gis.org/

172. AtlasAI. (2021). Helping You Decide Where To Invest In Africa and South Asia. https://www.atlasai.co/

173. iTech Mission. (2021). SDGs Global Dashboard. Explore, Monitor and Visualize SDGs Data. https://www.sdgsdashboard.org/

174. WHO/UNICEF. (2021). JMP Data. https://washdata.org/data

175. 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

176. 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.

177. 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.

178. 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).

179. Water.org. (2020). Women’s empowerment in water, sanitation & hygiene.

180. 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.

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

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

183. 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

184. 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.

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

186. 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

187. 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

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

189. Interview with Luis Andrés, Sector Leader and Lead Economist. (2020).

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

191. 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

192. 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.

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

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

195. 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

196. 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

197. 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

198. 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

199. 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

200. USAID. (2021). The DHS Program - Quality information to plan, monitor and improve population, health, and nutrition programs.

201. World Bank. (2021). Living Standards Measurement Study. https://www.worldbank.org/en/programs/lsms

202. Ghana Statistical Service. (2015). Ghana Poverty Mapping Report.

203. Tatem, A., Peter, G., Pezzulo, D. C., Weiss, D., & Bhatt, S. (2013). Development of Pilot High-Resolution Gridded Poverty Surfaces: Methods Working Paper.

204. Atlas: Environmental and Infrastructure. (2021). https://www.oneatlas.com/

205. Digital Globe. (2021). https://discover.digitalglobe.com/

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

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

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

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

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

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

aquaya

CONRAD N. HILTON
FOUNDATION