How urban sanitation data standards will improve data collection, sharing and analysis to accelerate and track progress on SDG 6.2.

John Peter Archer, Cofounder and Director, Gather (corresponding author)
Lindsey Noakes, Cofounder and Director, Gather
john@gatherhub.org
Geovation, Sutton Yard
65 Goswell Road
London
EC1V 7EN
United Kingdom

Abstract

Poor data practices across the urban sanitation sector have led to an unequal allocation of resources and an uncollaborative, ineffective sector. This has left the world unlikely to achieve Sustainable Development Goal 6.2. This paper outlines Gather's evidence-based recommendations for a sector-wide transformation in how data is collected, shared and used. Data standards play a critical part. The sector should be more targeted and specific in the data that it collects. This will reduce the cost and burden of data collection and analysis and will allow for prescriptive, geospatial analysis to inform areas for investment and collaboration.

Introduction

By 2030, 3.2 billion people will be living in cities without access to proper sanitation. Over the past 10 years there has been significant investment in new toilet technologies to meet the needs of underserved communities. These new innovations often do not need sewers, reduce water usage or reclaim valuable nutrients from human waste. Despite these advances, there has been no change in the proportion of people globally without access to adequate sanitation.

Gather was launched in 2016 as a research project to understand what was preventing the expansion of urban sanitation services in low income communities. We conducted interviews with nearly 100 professionals across a broad range of organisations providing sanitation services across east Africa and south-east Asia. Our research revealed that a lack of quality data on urban sanitation was making the sector less strategic, efficient and collaborative than it could be. We also noticed an acute need for the sector to better track its progress in providing sanitation for people living in cities and achieving Sustainable Development Goal 6.2: by 2030, achieve access to adequate and equitable sanitation and hygiene for all and end open defecation, paying special attention to the needs of women and girls and those in vulnerable situations.

Throughout our interviews we discovered that organisations providing sanitation services in low-income communities were collecting large quantities of data on their own operations, but that there was no mechanism to share and analyse this data to better understand the current state of sanitation across cities. We concluded that a sector-wide response needs to include behaviour change in the collection, sharing and use of data. Our conclusions were corroborated by the World Bank and United Nations. In 2017, the World Bank reported that poor data had led to an ineffective allocation of resources across

the sanitation sector.[1] In 2018, the United Nations found that the current lack of quality data had limited the sector's ability to monitor its progress and of "service providers, governments and development partners to be held accountable". It further recommended that "reliable, consistent and, whenever possible, disaggregated data are essential to stimulate political commitment, inform policymaking and decision-making, and trigger well-placed investments towards health, environment and economic gains."[2]

In response to our findings from our interviews, we tested three different approaches to learn more about how we can transform the sector's practices on data collection, sharing and analysis.

In 2017, we explored how we could help to improve access to urban sanitation data and geospatial analysis. First, we mapped shared sanitation infrastructure in the informal settlement of Mathare in Nairobi, Kenya. We wanted to understand the costs and challenges of collecting, curating and analysing sanitation data. Second, we focused on the creation of a data sharing platform. We wanted to understand what was needed for sanitation providers to share the data they were already collecting, so that it could be analysed and the insights shared with sanitation providers and funders. Both approaches revealed the need for a sector-wide data standard for urban sanitation data.

In 2018 we held the world's first data dive for urban sanitation. The purpose of the data dive was to explore how a data standard could ease the burden and cost of data collection on sanitation providers by simplifying the data they need to collect and share, and reducing the duplication of efforts. Our learnings from the data dive have informed the design of our new city data hubs. These hubs will create the virtual space for organisations to share standardised data and then access actionable, reliable insight that identifies where collaboration and investment is needed to expand sanitation services. The hubs are critical for key stakeholders in the provision of urban sanitation to gain the greatest value from standardised data. This will allow the sector to accelerate and track progress toward achieving Sustainable Development Goal 6.2.

Methodology

Our methodology is built on a collaborative approach to systems change. As well as continuing to listen to the experiences and needs of the sanitation providers we had interviewed, we also want to engage with experienced partners in the fields of data collection and data sharing.

In August 2017, we partnered with Spatial Collective – a local, respected community mapping organisation – to collect data on shared sanitation in the informal settlement of Mathare in Nairobi, Kenya. Mathare is home to an estimated 180,000 people. A team of eight young local enumerators spent two weeks collecting the GPS coordinates and 17 additional attributes for all 119 communal toilets in Mathare. We then analysed this data set alongside openly accessible data on population density from WorldPop at the University of Southampton.

---

[1] World Bank, *Reducing Inequalities in Water Supply, Sanitation, and Hygiene in the Era of the Sustainable Development Goals: Synthesis Report of the WASH Poverty Diagnostic Initiative* (Washington, DC: World Bank Group, 2017), xvi.

[2] United Nations, Sustainable *Development Goal 6: Synthesis Report 2018 on Water and Sanitation – Executive Summary* (NYC: United Nations, 2018), 9.

In June 2017, we began to design the architecture for an urban sanitation data platform with the support of Geovation, the UK Ordnance Survey's location data hub. A platform enables value-creating interactions between two groups of users.

In March 2018 we invited 21 data scientists to join us for a two-day data dive. Participants came from a broad range of universities, public bodies and businesses including: the London School of Hygiene and Tropical Medicine; University College London; the University of Leeds; Reed; Southward Council; KPMG; and Vitol. Half of the participants were female and one third of participants were from minority backgrounds; significantly above the industry average for data dive participation.

As of August 2018, we have started to invite members to join the first city data hubs for emerging cities in southern Africa.

Discovering the need for data standards: lessons from approaches to improving the availability of data by mapping sanitation infrastructure and improving access to data analysis by building a data sharing platform

We faced two challenges when considering how to improve the availability of data by mapping sanitation infrastructure. The first was the limited published guidelines on what indicators should be collected and how. Additionally, we faced poor economies of scale if we wanted to expand our survey to cover the entirety of Nairobi or to replicate the model in other cities inside or outside of Kenya. Our interviews with sanitation providers showed us that these challenges were shared across the sector.

Traditional data collection methods, such as household surveys, can be helpful for snapshot assessments. For example, our analysis revealed that only 3% of the population of Mathare had access to basic shared sanitation. However, we quickly discovered the limits of this approach for strategic decision making: the data quickly goes out of date, and regularly updating it is prohibitively expensive.

We were able to perform some basic analysis on shared sanitation in Mathare from the data that we had collected. To understand the state of sanitation across the whole of Nairobi though, we needed to supplement our data collection with data from sanitation providers working in the city. These organisations also wanted to access city-wide data sets, but there was no system available that allowed them to do this. We concluded that we needed to focus on improving the ability of sanitation providers to share their data and access analysis. We pivoted our focus to look at the creation of a data sharing platform.

In the early designs of our platform, we envisioned that organisations providing sanitation would upload data on their services. Gather would then perform geospatial analysis on this data and publish it as an interactive map. Finally, sanitation organisations would then use the map to make decisions on which parts of the sanitation value chain in a city they should prioritise for investment.

We started by launching two demos of our platform, showcasing the data we collected in Mathare in partnership with Spatial Collective. The first demo visualised the sanitation data and allowed the user to interact with it, e.g. display toilets connected to a sewer with handwashing facilities. The second demo visualised analysis we had performed on the

sanitation data alongside other data sets including population density. For example, users could make the map display where new toilets should be built, or where hand washing facilities should be installed. The purpose of these demos was to demonstrate our concept to sanitation providers so that we could work with them to design a fully operational platform where they could easily share and learn from data.

One early success of the launch of these demos was the interest we received from organisations providing sanitation. They offered us data sets on sanitation services in five cities to contribute to the building of our full working platform. We had previously assumed that these organisations would only share their data with us if major funders of sanitation were also signed up to the platform. We had underestimated the strength of another incentive: these organisations wanted to understand what their data could tell them, but often did not know what questions to ask their data or how to ask them.

When we designed the demos, we had assumed the technology would be a significant challenge. Our workshops with Geovation, and a project designing a crisis response tool with IBM Research, taught us that readily available geospatial software and programme architecture could perform the analysis we required. However, when we received the first data sets from sanitation providers, we discovered that in fact our biggest challenge was the quality of the data that we had been sent. There was little uniformity across the data sets, apart from the GPS coordinates of the toilets. Furthermore, the surveys had frequently been designed with limited consideration of the analysis that would need to be performed on the data. We could perform exploratory analysis on the data sets (like the analysis on our data from Mathare) but we were unable to perform deeper analysis on the individual data sets or perform comparative analysis between them.

The lack of standardisation across the data we received meant it would be necessary for us to recommend a new data standard for the collection and management of urban sanitation data before we could proceed with building a data sharing platform. Data standards are crucial. It is very difficult to properly compare or collate data sets when the data has been captured in different formats through different surveys that have relied on different questions and definitions. We can see from the data sets that were shared with us that the sanitation sector predominately collects data on operations. This focus can be useful in understanding an individual organisation's operations. However, it does not allow the sector to share data and understand the level of sanitation in a city. This makes providing sanitation in a strategic, collaborative way much more difficult.

Further demonstrating the need for data standards: hosting the world's first data dive for urban sanitation

In March 2018, we hosted a data dive to explore the benefits of a new data standard for urban sanitation. 21 data scientists analysed the data that sanitation providers had shared with us. We were originally going to use five datasets for the event to compare data from different urban sanitation providers. However, the quality of each dataset was far lower than we had expected, and none came with codebooks. We decided to refine the focus of the dive to look at just one comprehensive dataset from one urban sanitation provider.

The dataset we used for the data dive came from an urban sanitation organisation working in the informal settlement of Kanyama in Lusaka, Zambia. Kanyama was established as an unplanned settlement in the 1960s and has grown to become the largest informal

settlement in Zambia. Estimates of its population range from 143,000[3] to 370,000.[4] People in Kanyama are heavily reliant on pit latrines for their sanitation needs, with around 80% of people using them.[5] Kanyama frequently experiences severe flooding.

The dataset is the largest available sanitation dataset of this type within the sector. It is a census-type dataset that surveyed each household within a defined area and had a coverage of more than 100,000 people. It includes data for a wide range of variables, including the type and state of toilet facilities, and current faecal sludge management practices.

The data scientists investigated the validity and reliability of the data, how the data could be used to estimate the coverage of sanitation services and assess health risks to the population, and how the data could be used to optimise waste collection routes. The data scientists were able to create geospatial visualisations of the data to provide insight, but the variable quality, format and structure of the data confirmed the need for a sector-wide strategy for data collection and management.

The key findings from the data dive were:

- Large-scale, one-off surveys produce are useful for descriptive analysis to gain a basic understanding of the state of sanitation in an area.
- Large-scale, one-off surveys are not as useful for prescriptive analysis for what could and should be done to improve sanitation in an area.
- The data set included questions that are not objectively assessable.
- Microsoft Excel should not be used for data management, sharing or analysis because it is more difficult to track edits, formatting and cleaning.
- Standardised data should be collected regularly over time to enable the development of algorithms to provide improved, prescriptive analysis that could help improve route optimisation and investment into other parts of sanitation value chain.
- Standardised data would contribute to more accurate risk scores to understand an area's vulnerability in times of public crisis.
- Standardised data would allow for the creation of standardised zoning of informal settlements.

Applying data standards to improve collaboration and efficiency across the sanitation sector

Our research has shown that sanitation organisations have spent resources on high cost, one-off surveys that provide a snapshot of the state of sanitation, but then quickly go out of date and are not designed for predictive or prescriptive analytics. Therefore, we recommend that the urban sanitation sector should stop collecting lots of low-quality data and should instead focus their data collection to produce smaller quantities of higher quality data. They should also make use of available complementary open data sets.

[3] Lusaka City Council, 2016
[4] UNICEF, Zambia Cholera Report #2 (Lusaka: UNICEF, 2016), 1.
[5] WSUP, *Development of a Toilet Database For Kanyama Peri-Urban Area, Lusaka* (Lusaka: WSUP, 2017), 7.

We also recommend that the sector should move away from linear data collections and focus on embedded, ongoing data collections. Data collections should be standardised, focus on what can be most objectively assessed and be regularly undertaken as part of existing activities. If standardised, this data then be shared between organisations to improve the quality and usefulness of the analysis.

We applied this recommendation when we coached a team for Geovation's second annual Crisis Hack in April 2018. A team of four consultants spent three days developing an SMS-based tool that embedded data collection in waste collection services. The purpose of this data collection was to optimise toilet waste collection routes. Data would be collected by businesses providing toilet emptying services; after a toilet had been emptied several times and data collected each time, it would be possible to perform further predictive analysis and gain guidance on how best to improve sanitation services in the area.

To apply these recommendations, we are pursuing the creation of a data standard in a similar manner to the creation of the geospatial standards for GeoJSON and IETF. The aim will be to assess what is currently being collected by organisations and then refine and standardised ten key indicators on the sanitation value chain. We anticipate that these indicators will include: the location of the infrastructure; the amount of waste contained; the amount of contained waste that is transported; and the amount of transported waste that is treated.

We are also pursuing the creation of data hubs that reduce the cost and burden of data collection and analysis and allow for prescriptive analysis to inform areas for investment and collaboration. These hubs will launch with three key stakeholders within a specific city: a local government statistics office, a container-based sanitation provider and a utility-backed sanitation provider. Once formed, hub members will agree to implement the new data standard and recommended data practices into their ongoing data behaviours. They will then be able to share their standardised data for aggregated, geospatial analysis that will provide reliable, actionable insight that insight that identifies where collaboration and investment is needed to expand sanitation services. Data hubs will create the virtual space for organisations to collaborate and make data-driven decisions.

Conclusion

The urban sanitation sector needs to adopt a sector-wide data strategy, and this starts with the adoption of a data standard for urban sanitation. A data standard would ease the burden and cost of data collection on sanitation providers by simplifying the data they need to collect and share and reducing duplication of efforts. Standardised data, accessible through a data sharing platform at the centre of a city data hub, will allow organisations across the urban sanitation sector to work more effectively, efficiently and collaboratively to provide sanitation to people living in low income communities. This approach will help the sector work toward achieving Sustainable Development Goal 6.2

Acknowledgements

The authors would like to thank our data collection partners at Spatial Collective, the sanitation providers who have shared their datasets with us and the 21 data scientists who contributed their time, expertise and enthusiasm. We would also like to thank Gather's

panel of expert advisors for their strategic input as we seek to contribute to the vital work of the urban sanitation sector.

Bibliography

Lusaka City Council, Lusaka, Zambia, 2016.

UNICEF, *Zambia Cholera Report #2*. Lusaka: UNICEF, 2016.

United Nations. *Sustainable Development Goal 6: Synthesis Report 2018 on Water and Sanitation – Executive Summary.* NYC: United Nations, 2018.

WSUP, 2018. *Development of a Toilet Database For Kanyama Peri-Urban Area, Lusaka.* Lusaka: WSUP, 2017.

World Bank, *Reducing Inequalities in Water Supply, Sanitation, and Hygiene in the era of the Sustainable Development Goals: Synthesis Report of the WASH Poverty Diagnostic Initiative.* Washington, DC: World Bank Group, 2017.