  Journal of Water, Sanitation and Hygiene for Development | 07.4 | 2017

Research Paper

Why do water points fail? Learning from open-ended failure descriptions in the WPDx dataset

Joseph Cook and Sheena Lahren

ABSTRACT

Despite the explosion of geospatial data collected on water points, there have been relatively few efforts to date to use these data to understand the correlates and potential causes of water point breakdowns. We add to this literature by coding open-ended responses around functionality status in the open-access WPDx (Water Point Data Exchange) database. We code responses into 41 different categories of functionality problems for 244,075 water points from 25 countries in Africa, Asia and Central America, though we narrow our analysis to 41,716 points in six countries. We find that descriptions of technical breakdowns or problems are most common, though concerns about water resources and water quality feature prominently. We also find that 7% of records in our analysis subset spontaneously mention concerns over vandalism or theft, something which has received relatively little attention in the sector. Information on user financial contributions is mostly omitted by uploaders to WPDx, but we find that users are not contributing at all in 40% of the 96,651 water points with these data. Our results should be interpreted cautiously considering the obvious selection and subjectivity concerns, but the analysis highlights the potential benefits of coordinating an augmented data collection standard.

Key words | functionality, management, water point

Joseph Cook (corresponding author)
School of Economic Sciences,
Washington State University,
Pullman, WA,
USA
E-mail: *jhcook@uw.edu*

Sheena Lahren
Lively Minds,
Tamale,
Ghana

INTRODUCTION

By one estimate, one-third of rural handpumps in sub-Saharan Africa are not functioning at any given time (RWSN 2010). This comes despite the approaches many governments, non-governmental organizations (NGOs), and donors have taken to ensure sustainability through emphasizing community management and demand responsive approaches (McPherson & McGarry 1987) and creating standards for construction (RWSN 2013). According to one estimate, water supply failure in Africa 'represent[s] a lost investment in excess of $1.2 billion' (Baumann 2009). Some studies have, however, found more encouraging functionality rates: 79% of 1,509 water sources were functioning in the Afram plains of Ghana (Fisher *et al.* 2015) and 95% of sources produced water in 400 rural communities in Bolivia, Peru and Ghana approximately seven years post-construction (Whittington *et al.* 2009).

Researchers and WASH professionals have long been interested in understanding the underlying reasons for why some water points are maintained and continue to serve through their full design life while others break down but are not repaired. In this study, we define water points as improved water sources (as per the standard Joint Monitoring Programme definition), which include: boreholes with handpumps, water points on gravity-fed systems, protected dug wells, standpipes, and protected springs. From an engineering perspective, one might be interested in the immediate, technical causes of failure, such as a water source drying, a gravity system main breaking, or a

doi: 10.2166/washdev.2017.022

submersible pump failing. Going further, one may be interested in whether those technical failures are more common in some projects because of poor construction quality (including the use of cost-cutting but inferior materials), unmonitored construction methods, and inappropriate siting (Harvey 2004; Danert *et al.* 2010; Danert 2013). Underlying these questions, however, is why the water point was not maintained to prevent failure or promptly repaired. These reasons include poor community participation and low user demand (McPherson & McGarry 1987; Isham *et al.* 1995); lack of capacity of communities to maintain handpumps through the inability to raise the necessary finances or difficulty in accessing spare parts (Harvey *et al.* 2002; Franceys & Pezon 2010; Schweitzer & Mihelcic 2012); and the lack of involvement or failure of government and the public sector (RWSN 2014; U-Dominic *et al.* 2014).

The ability of organizations to collect georeferenced data on water points has increased dramatically with the widespread adoption of low-cost smartphones and the development of software tools such as AkvoFlow to upload, synchronize and display these data. As a result, the rural water supply sector has bolstered efforts for consistent water point mapping, monitoring standards, and methodologies, including frequent updating of water point status and sharing of data (RWSN 2014). The sector has had a mixed record, however, of harnessing and aggregating these data collection efforts for rigorous empirical research into correlates of sustainability. There is a risk that each NGO or delivery provider keeps the underlying data it has collected within its own domain. It may perhaps glean actionable lessons from its in-house analysis of that data (arguably the most important outcome of the exercise), and it can display attractive maps of its projects for donors or elected officials. However, the data and analysis risk sitting on the proverbial shelf, like so many of the (analog) monitoring and evaluation reports of the past decades.

Three notable studies are exceptions and cause for optimism in the use of 'big data' in service of understanding water point functionality. Foster (2013) analyzed operational, technical, institutional, financial, and environmental predictors of functionality for data collected from over 25,000 community-managed handpumps in Liberia, Sierra Leone, and Uganda (one of the largest available datasets at the time of the study). The study found that the risk factors associated with non-functional handpumps in all three countries were water point age, distance from the city, and lack of a user fee. Fisher *et al.* (2015), using data collected from 1,509 water points in the Greater Afram Plains Region of Ghana, distinguished among: (a) determinants that affect borehole failures that were somewhat difficult to control (e.g. number of water sources per community, number of users, hydrogeological variables); (b) those which were modifiable by water point management (e.g. tariff collection); and (c) those modifiable by implementers and local agencies (e.g. access to spare parts, mechanics). Bonsor *et al.* (2015) investigated a much smaller number of water points ($n = 24$ as a pilot, in northern Uganda) than Foster (2013) and Fisher *et al.* (2015), and only those that were abandoned, but used far more comprehensive data collection methods that enabled the researchers to look beyond factors associated with failures to the underlying root causes. They found poor siting and construction of water points as the most significant determinants of water point failure. The study went deeper to identify that these issues are likely due to an absence of expert supervision (hydrogeologists and engineers), which was influenced by poor procurement practices, shortage of trained staff, and lack of knowledge of the importance of these factors for successful water points.

In this study, we use the largest existing downloadable dataset of water points, the Water Point Data Exchange (WPDx), to shed light on reasons why water points fail. Despite the broad use of the WPDx platform, we know of no study examining or categorizing the open-ended responses on functionality status. As we discuss below, this dataset has advantages as well as limitations. Our intent in this article is not to make definitive causal claims on which factors lead to water points providing sustained service. Rather, one can view this study as an inductive, bottom-up approach for what types of factors are spontaneously mentioned by respondents when assessing whether water points are functioning. These factors will include both immediate technical failure reasons as well as underlying factors contributing to poor maintenance. This type of analysis can complement studies such as Foster (2013), Fisher *et al.* (2015) and Whittington *et al.* (2009).

DATA

The WPDx was established in 2015 to complement improved data collection efforts and provide an open-access platform in which actors in the water sector can upload and access water point data worldwide. The aim of WPDx is to harmonize the various data sources becoming increasingly available to create a better understanding of the status of water services. To ease the process of sharing data, the criteria for uploading data are minimal. The uploader need only include four core attributes – the location, the organization which collected the data, the date the data was collected, and functionality status of each water point (specifically, is water present when assessed: 'Yes,' 'No,' or 'Unknown'); see Appendix Figure S1 for a screenshot of the upload process (available with the online version of this paper). The uploader must also include a description of either the water source or the water point technology. Optional information that can be uploaded as open-ended responses includes management type, installer, payment, installation year, and an open-ended text field ('*status*') providing 'any descriptive status regarding the condition of the water point' (see https://www.waterpointdata.org/sites/default/files/page/wpdx_hashtags_0.pdf). We know of no explicit incentive for uploaders to share more than the minimum required fields.

We downloaded the WPDx dataset on 24 December 2015, at which point the database contained records on 244,075 water points from 25 countries in Africa, Asia and Central America. WPDx contains several large datasets that have been previously released publicly and used in prior research. We suspect there is substantial overlap with Foster (2013), who analyzed approximately 25,000 handpumps in Liberia, Sierra Leone and Uganda. Nevertheless, it includes data that we believe have not been previously analyzed, including 59,850 water points in Afghanistan, the majority implemented by DACAAR, the Danish relief agency. As of November 2016, the dataset had grown to 294,229 water points.

Before describing our approach for coding the open-ended *status* field below, we describe some of the basic characteristics of the WPDx dataset (again, as of December 2015). Because data come from organizations that self-select into uploading to WPDx, these statistics cannot be said to be representative of all rural water points. Seventy-four percent of points are from sub-Saharan Africa (Appendix Table S1, available online), predominately from Uganda, with most of the remainder from Afghanistan. Coverage in Central and South America is limited ($n = 1{,}292$ points, 0.5%).

The type of management is reported in 48% of records. Among those, the large majority (78%) are community-managed. Nine percent are managed by a private or 'delegated' entity, 9% are 'institutional management', and 1.5% are managed by the government.

Ninety-two percent of records recorded the type of technology, though not using standardized categories. We coded 491 unique text fields into groupings: 45% were handpumps or drilled boreholes, 22% were shallow or hand-dug wells, 6% were gravity-fed systems, 11% were springs (protected and unprotected), 7% were rainwater collection systems, $<1\%$ were surface water sources, 5% were called 'standpipes', 'standposts' or 'kiosk' and could refer to a number of different types of systems, and 3% could not be classified based on vague descriptions.

A useful installation year was given in 88% of records; the report date was provided in 99%. The median age of water points (calculated as report date – installation year, with some data cleaning) was five years. Half of the 213,320 points with age data were five years old or younger; 76% were less than 10 years old (Appendix Figure S2, available online).

Based on the 'status_id' variable, 74% of water points in the dataset were functional (i.e. producing water) at the time of the survey (Table S1). We again caution that readers should not take these figures to be representative of water services in these countries; the Joint Monitoring Program provides the most current and comprehensive country-level estimates of access to improved drinking water sources (though not functionality of individual water points). Our aim here is not to weigh into the debate about how to define functionality (see Carter & Ross 2016), and one would expect a certain fraction of water points to be down temporarily for repairs or rehabilitation. There are no data systematically provided in WPDx to distinguish these from permanently abandoned water points, though the open-ended *status* field does sometimes give clues in this regard, as we describe below. Functionality status was either missing or marked as 'don't know' in 1.82% of records, though these are concentrated in just a handful of countries (see Table S1). Appendix Figure S3 (available online) plots the fraction of water points still functioning by system age, showing a clear decline over the first 10 years (as would be expected, again see Carter & Ross (2016)). It then shows an increase in functionality as systems are (presumably) rehabilitated in their second decade, and then a noisier pattern as the number of systems older than 20 years declines and the confidence intervals widen substantially as the sample size in each age bin declines.

In many cases we found that the open-ended *status* field indicated problems with a water point that was labeled in *status_id* as being functional. After coding each text entry in *status* by hand (described next), we find that 15.8% of sources listed as functioning have some sort of qualification or issue mentioned; for example: 'Poor water quality', 'water point concrete problem' and 'working but with problems – pump broken'. A small (~0.2%) number of these text fields confusingly suggest that the source is in fact not working, such as 'Functional – it's totally broken down without repair', 'Functional – borehole not functional', and 'Functional – water is not coming'. This confusion might be explained by whether 'functionality' was referring to the extraction technology, the water source, or both. For example, a handpump may itself be broken but it might still be possible to fetch water with a rope and bucket. Conversely, the water source itself is dry (perhaps temporarily during the dry season) but the handpump is still mechanically functional (we thank a reviewer for pointing out this possibility).

The installer was mentioned in 51% of records, though in 6,276 unique text codes which we did not attempt to parse. Nevertheless, 52% of these records are from five implementer text codes: The Danish Committee for Aid to Afghan Refugees (DACAAR, 31% of these records, all in Afghanistan), the Centers for Disease Control and Prevention (CDC, 7.7%), 'Private person' (6.5%), WorldVision (3.7%), and 'Government' (3.2%).

Our original intent in downloading the WPDx dataset was in fact to try to measure across countries the prevalence, type, and level of user fees using the '*pay*' field. We learned, however, that the field is not commonly reported among those uploading data to WPDx: this field is missing in 56% of records and another 10,348 say 'not recorded' or 'no water'. Among the 96,651 remaining records, 52% imply that users regularly contribute in some way, 40% that they never pay, and 8% that they pay only when the system breaks down. Among the records indicating some regular user contribution, 40,918 (82%) report no further useful information (i.e. 'Yes'), or report information which may or may not indicate that contributions are actually being collected (i.e. 'Water committee collects fees', see Carter *et al.* 2010). Among the remaining 9,051 records, 35% are volumetric (per jerrican, per animal, or per m^3) and 65% are periodic, most commonly per month.

CODING METHODOLOGY

The major task of this study was to develop categories to describe reasons for failure mentioned in the open response data in the WPDx dataset (the *status* variable) and then code these variables for each of the unique open responses. We first used existing studies, particularly Bonsor *et al.* (2015), to construct a typology of different types of failure causes (Figure 1). These include broader categories such as 'management issues' as well as subcategories such as 'No Mechanic' (open response mentioned a lack of mechanics as contributing to the failure) and 'Funds for Maintenance'. We next used Stata 14 to collapse the open-ended responses in *status* among the 213,846 water point records (of 244,075 in total) that had some information in this field. This produced 13,609 unique text entries, which were copied into a spreadsheet for ease of coding. For example, the WPDx data contains 234 water point records where status is recorded as 'Always functional' and 12 where status is 'No – broken down. fuel'; these are two of the 13,609 text codes in the spreadsheet. Appendix Table S2 (available with the online version of this paper) provides more examples of text codes, and the full spreadsheet containing all unique text codes as well as our codings is available at http://bit.ly/2jbFaeF.

The second author assigned codes to each open-ended response in the WPDx data: each small circle in Figure 1 is a dummy variable equal to one if the responses mentioned that issue and 0 otherwise. The first coder revised these categories as coding progressed. Appendix Table S3 (available online) provides further description of each of the codes. To guard against idiosyncratic interpretations of text, the first author then independently coded each entry using the final coding system, making only a few additional changes to the coding scheme. We then cross-checked the differences in coding decisions and agreed upon which of the two codings to use in the analysis. We then merged this final coded dataset with the full WPDx data for the analysis below.

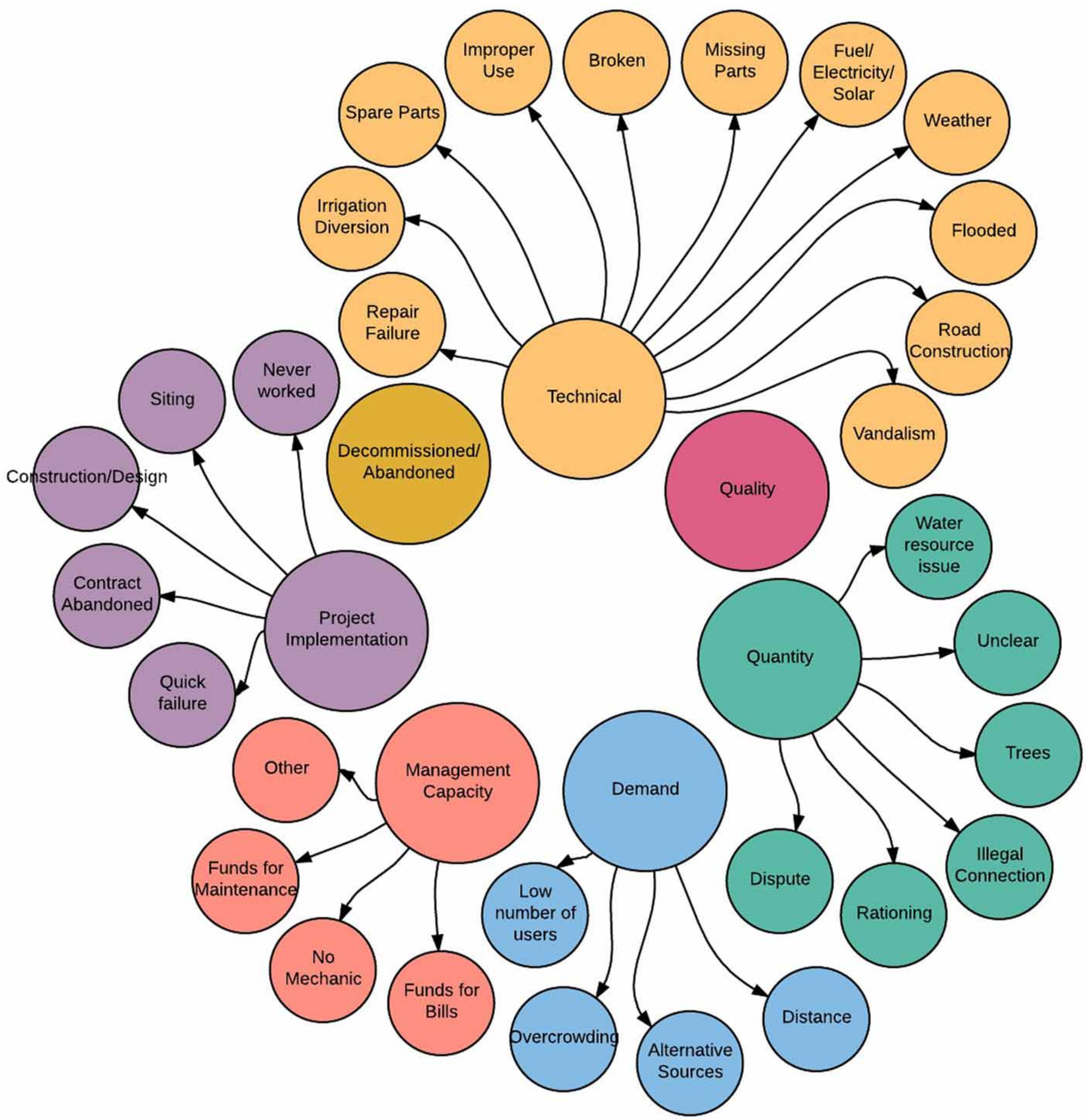

Figure 1 | Schematic of data coding (see Appendix Table S3 for more description of each category, available with the online version of this paper).

A given text entry could be coded for multiple causes simultaneously. For example, the entry 'No – broken down. Well damaged | Well polluted | Apron damaged | No cover slab' is coded as both *Broken* and *Missing* within the Technical Failure category and as a *Quality* issue. We also tally the number of codes identified for each *status* entry.

RESULTS

Our focus is on learning about underlying reasons why water points fail or are in danger of failing, so we begin by describing, and then discarding, codes which are not helpful for that purpose. (Appendix Figure S4 (available with the online version of this paper) provides a visual depiction of this exclusion process.) First, of the 244,042 entries in the WPDx data, 12% of water points provide no information on *status*; 88% of data uploaded in the database contained something in this field. Second, 129,064 entries (53% of the total) are functional water points that provide no qualifier or any indication of a problem. Examples of these would be a *status* entry reading ‘Functional’ ($n = 72{,}728$), ‘Yes – functional’ ($n = 17{,}939$), and ‘Normally operational’ ($n = 276$). Third, 21,096 entries are for non-functioning water points, but *status* provides no useful information on why the point is not functioning. Examples include ‘LOST’, ‘Non-functional’, ‘Problem but not delivering water’ and ‘non-operational’. We also include in this category the entry ‘Dry’ ($n = 3{,}708$), which we interpret to mean the point is not producing water, though it could plausibly be interpreted as a functional water point that is only intermittently delivering water.

To further narrow our focus, we next exclude points that have only one entry and where that entry is either *Vague* ($n = 12{,}991$), *Decommissioned* ($n = 726$), *UnderConstruction* ($n = 193$), or *Abandoned* ($n = 375$). In other words, we drop records where there is little useful information about why a water point might be failing or have failed. We also drop 61 records that are combinations of these (i.e. a *status* reading ‘Decommissioned – abandoned’).

Next, the survey effort in Tanzania seems to have used an electronic data collection form that recorded the year a system broke down (a typical entry is ‘Status:Functional|Breakdown Year:1985|Quantity:Enough|Quality:Soft’). It appears that the survey attempted to census all water points, even those that have long since stopped working. It is also possible that this simply records when the last known breakdown occurred. To be conservative, since the information from such a long-dormant source would likely be unreliable, we drop 5,710 records that have a breakdown year, all of which are 2008 or earlier (for reasons we do not know).

This leaves 41,716 records, or 17% of the total records, all of which now focus on water points that are either ‘not functional’ or were listed as functional but for which we found qualifiers embedded in the *status* field. Ninety-nine percent are from six countries: Tanzania (22% of the remaining records), Afghanistan (25%), Uganda (22%), Sierra Leone (21%), Liberia (7%) and Kenya (2%). Of the remaining records, we interpreted the *status* field as triggering only one variable in 79% of water points; 16% triggered two variables and 4% triggered three. Less than 1% triggered four or five codes.

Not surprisingly, the most common thing enumerators mention in *status* is a technical description of something being broken. Examples include ‘Tap broke down’, ‘technical breakdown’, ‘pipes stuck at bottom’, ‘leaking pipes’ and ‘apron damaged’. This code (*Broken*) appears in 26,064 (62%) of the remaining entries. In 20,244 of these entries, it is the only code triggered. In other words, the enumerator provided no additional information about non-technical reasons why the water point is failing or failed, although it is certainly plausible that other underlying conditions led to this breakdown. Since our objective is not to distinguish between types of technical problems, we drop these 20,244 records where the only code triggered is *Broken* and we learn little about any underlying issues. This leaves 21,472 records, of which ninety-percent are from three countries (Tanzania 41%, Uganda 28% and Sierra Leone 21%). The remainder are Afghanistan (2.8%), Colombia (0.07%), Guatemala (0.07%), Honduras (0.03%), Kenya (3.3%), Liberia (3.0%), Mexico (0.4%), and Nigeria (0.2%). Nearly all the records in the WPDx dataset from Afghanistan listed only a technical problem in *status*. Among these, 59% triggered one code, 31% triggered two codes, and 10% in total triggered three, four or five codes.

Among this subset, quantity and quality concerns feature more prominently along with other technical problems (Table 1, column ‘All’). We also provide (in Appendix Table S3) the raw tabulations for each of the variables in Figure 1 for the entire WPDx data with non-missing *status* fields. We provide the correlation matrix of the groupings in Table 1 in Appendix Table S4 (available online). Most correlation coefficients are negative and statistically significant at the 95% confidence level, but with modest magnitudes of 4–20%. Quantity and Technical (except Broken) are more strongly and inversely correlated, as are Broken and Quantity, suggesting that some technical problems could have been expressed using language we coded as a quantity (e.g. ‘yield low’ or ‘low pressure’). Not surprisingly, Broken and Technical (except Broken) are strongly and positively correlated: technical descriptions of why a point was not working may have included multiple types of problems (i.e. ‘broken handle, valve missing’). Because of the way we

Table 1 | Frequency of failure types

Category	All Freq (% total)	Handpumps and boreholes Freq (% total)	Points on gravity systems Freq (% total)
Project implementation	856 (4%)	242 (3%)	20 (1%)
Quantity	10,025 (47%)	4,599 (56%)	2,925 (85%)
Quality	3,893 (18%)	2,187 (27%)	223 (7%)
Technical (exc. Broken)	6,623 (31%)	1,311 (16%)	209 (6%)
Broken	5,820 (27%)	1,098 (13%)	173 (5%)
Demand	426 (2%)	147 (2%)	73 (2%)
Management	1,217 (6%)	428 (5%)	114 (3%)
$n =$	21,472	8,206	3,172

Note: See Figure 1 and Appendix Table S3 for further description of categories (Appendix Table S3 is available with the online version of this paper). Because a given *status* entry could include multiple codes, the second column sums to more than the total sample and the third column sums to more than 100%. The remaining types of water systems not reported in the second and third columns are springs ($n = 1{,}082$), piped systems ($n = 251$), rainwater systems ($n = 675$), surface sources ($n = 34$), standpipes or standposts ($n = 1{,}039$) and water points where the technology was missing or could not be determined from the description or data given ($n = 4{,}787$).

defined the *Quantity* variable, many of the most commonly cited *Quantity* problems could also be technical problems: 39% of these records are from a *status* entry 'Status:Functional|Quantity:Insufficient|Quality:Soft' and 19% are from 'Status:Not functional|Quantity:Dry|Quality:Soft'. We also coded a substantial amount ($n = 2{,}081$, or 9.7%) of these *Quantity* problems as *Water Resource* problems that may be related to seasonal droughts or intermittent water sources. Note that Appendix Table S3 lists *Water Resources* as being mentioned in 2,475 water point *status* entries. This is for the entire WPDx dataset, including Tanzanian records that included 'breakdown years' and were dropped, as described above. This is true for all the codes. Some common entries coded as *Water Resource* include any mention of the source being seasonal ('In the dry season'), and phrases like 'the spring has dried up'. The word 'table' was found in 142 water point records. Again, however, phrases like 'well is dry' could connote issues with the water table or with a technical malfunction of the pump. Finally, recall that our analysis excluded records tagged as 'Dry' ($n = 3{,}708$) because of ambiguity. Were we to include these and code them as a quantity issue (a plausible interpretation), the percentage of records for Quantity would rise from 47 to 54.5% ($10{,}025 + 3{,}708/21{,}472 + 3{,}708$). Readers who wish to interpret 'Dry' as in the 'Technical' or 'Broken' categories can similarly calculate revised percentages.

Quality issues were mentioned in 18% of the remaining records, with common phrases including 'well polluted', 'quality:salty', 'quality:milky', and 'dirty water'. In a third of these records there was also some indication of something being broken or another technical problem.

Project implementation problems featured in 4% of the remaining records. Seventy percent of these mentioned problems in Construction and Design, with phrases such as 'poor installation', 'pipe short', 'needs protection', 'defective pump parts', and 'poor workmanship'. Siting problems were mentioned in 146 records (17% of the *ProjectImpl* group), and just under 100 water points seem to have either never worked or failed shortly after construction.

Demand concerns were raised in only about 2% of records ($n = 426$ records). Distance to the source and possible overcrowding problems were rarely mentioned. The more common of these problems were either low demand from users ($n = 96$) with phrases like 'Functional – no demand', 'it is ignored completely', 'no water users', and 'lost interest due to the poor yield' (also coded with *Quantity*, above), or low demand specifically because alternate sources are available ($n = 253$). Phrases for the latter include 'cheaper alternative sources', 'free (unimproved) sources', and 'people are connected to piped water', suggesting that a water point was abandoned because service was upgraded, a good thing.

Management concerns arose in 6% of records. The two most common subsets of these concerns were a lack of financial capacity for operations and maintenance ($n = 522$, 43%) and an overall lack of community capacity to manage the source ($n = 591$, 49%). Example phrases for these included 'lack of funds for repairs', 'communities failed to raise funds for repair', 'no funds for buying spares', 'committee are corrupt and failed to do their work', 'the water source committee not active', 'poor management of the scheme', and 'embezzlement of funds'. A lack of pump mechanics or technical ability to fix the pumps was mentioned in only 41 records.

Finally, technical concerns other than some indication that things were *Broken* were present in 31% of records. The most common issues in this subset was that something was missing (i.e. 'no pump', 'lack of gutters', $n = 2{,}535$) or that a source was either blocked or silted up ($n = 2{,}532$). We also found 116 records suggesting a source was broken because of improper handling by the users ('carelessness of community members', 'taps were broken by pupils', 'suspected of being misused by the community'). A lack of fuel or electricity to run the system was indicated for 176 water points, and missing solar panels were a relatively common reason for this lack of electricity. In fact, we were surprised by the frequency of mentions of vandalism or parts being stolen: 1,467 *status* entries mentioned this, or 6.8% of our subset of 21,472 records. Indeed, the words 'steal' or 'stole' occur in 1,188 records, and 'vandal', 'vandalism' or 'vandals' in 149 records. The most common phrases (about two-thirds) included 'No – broken down, pump stolen' and 'broken down system, pump stolen', indicating that perhaps the theft was an *ex-post* consequence, and not a cause, of the water point failing. However, a number of records indicate more malicious sounding situations like 'tank axed and tap stolen', 'stones thrown in the casing', and 'if tap heads are replaced, people come and steal them at night'. By country, we find mentions of vandalism more common in Liberia (40% of 639 records) than in Uganda and Sierra Leone (both 11%) and Tanzania (0.3%).

Our analysis conflates different types of water delivery technologies. The second and third columns of Table 1 show results for two types of system technologies that were somewhat clearly defined: handpumps or drilled boreholes, and gravity-fed systems. The proportion of records showing demand or management concerns is roughly the same as the overall sample. Handpumps are more likely to report *Quality* issues, and both handpumps and (especially) gravity systems were more likely to report *Quantity* concerns. The table also indicates that both handpumps and gravity systems were less likely to report something being *Broken* or other *Technical* problems than the full sample that includes wells, springs, etc. This is, however, among the sample of entries where there was at least one additional code triggered beyond *Broken*. Of the 20,706 handpumps that had useful *status* codes, 60% were coded with only *Broken* compared to 48% overall. The similar figure for gravity systems was 10%.

DISCUSSION

We begin by briefly discussing the limitations of the study. First, as noted earlier, organizations and governments self-select into sharing data with WPDx, so our data cannot be said to be representative of all water points in low-income countries. The selection issue should also bias functionality results upwards, since implementers with less successful programs are almost certainly less likely to want to publicize their failures.

Ideally the dataset would provide information on the data generating process for each entry, most importantly the survey questions. Even with that information, the selection and training of enumerators may lead to unobservable differences in how much they probe when asking these questions, given that water point failures likely stem from a complex chain of events. The WPDx dataset provides only the 'source' of the information, which we tabulate in Appendix Table S5 (available with the online version of this paper) for both the full dataset and the smaller analysis dataset. It is possible in some cases to link this 'source' field to one of the large mapping exercises for which more information is available online (e.g. WASHLiberia, 'WA', Ministry of Water and Environment Uganda; links to these initiatives are http://wash-liberia.org/; http://wpm.maji.go.tz/ (Tanzania mapping program); and http://wateruganda.com/), though not all have readily accessible survey instruments and documentation. Most text codes for 'source' do not, however, have an obvious link to a survey effort but simply refer to an organization, which may change its survey protocols year-to-year or country-by-country. Given the quality of the WPDx data and the potential for adding selection bias problems, we feel it is beyond the scope of this paper to attempt to seek out the exact survey questions for over a dozen 'source' codes and attempt to parse our analysis further. On a related point, we note that misinterpretations in the 'status' field could occur because of misspellings, improper grammar, language barriers (particularly in non-Anglophone countries), and vague descriptions by the data collector.

Third, the open response data are limited by discrepancies on definitions of functionality, typically measured as a binary, functional or not. For example, 'pipes leaking' is labeled as functional by some data sources and non-functional by others. This evidence aligns with a study by Carter & Ross (2016), which argues that functionality is open to interpretation, and different definitions of functionality used across the water supply sector inhibits the comparability of data. In addition, as noted earlier, the *status* entry of some functional water points clearly imply that the source is non-functional (i.e. 'Functional – all parts stolen.')

Finally and most critically, manually coding and interpreting the open response status data in the WPDx dataset is subjective. For example, 'the pipe was broken by users' could be interpreted as a malicious act of vandalism, as an accident, or as improper use. 'Lack of community contribution' may refer to lack of motivation or participation, but this description likely refers to lack of funds contributed by the community. In addition, descriptions of failure reasons often imply linkages to other failure reasons not described, forcing each coder to make a judgment. For example, 'Located in the swamp' clearly seems to be a *Siting* issue, but is likely also to be a *Quality* issue and could affect *Low Demand* among users. 'Broken and not repaired' is a technical issue, but could imply poor management, lack of funding, or access to spare parts issue. In this study, we coded *only* the direct description of failure reasons and did not 'read into' descriptions. Although there may be financial, operational capacity, management, or environmental factors that explain the failure types we see, the dataset does not contain enough structured information on these issues. However, the difficulty in even coding the direct descriptions is apparent in our modest inter-coder reliability measures (see Appendix Table S3): they average 0.58 (sd 0.24), and five of 36 variables have a correlation coefficient below 0.30.

Overall, then, what did we learn? Not surprisingly, descriptions of technical problems are by far the most common type of response given for non-functioning water points or water points that are functioning but with some qualification. If one sends an enumerator to census water points and instructs her that she should, upon finding a non-functioning point, tell us 'why is it not working', one would expect most enumerators would record the most obvious symptom they see. Many could also come from technical or engineering backgrounds, and we do not in any case know exactly what enumerators were asked to do or who they may have interviewed as part of their data collection. Still, glimpses from other comments, though relatively rare, show the complexity of managing water points. Concerns over water resource issues feature fairly prominently, and in many countries these concerns may grow with groundwater depletion and as hydrological patterns shift. On the other hand, whether or not an enumerator observed a water resource problem is likely affected by whether they are collecting data in the rainy or dry season. Some codes indicated that wells were permanently dry, but many were ambiguous: the water points may resume working when the rains arrive and recharge aquifers. A detailed analysis matching survey dates with rainy season dates in six different countries is beyond the scope of this paper.

Likewise, a large fraction of responses described problems with the taste or quality of water, which could make users reluctant to contribute to upkeep or even abandon sources. We were surprised at how infrequently responses included information on management or financial capacity, particularly how often the *pay* variable in WPDx was blank or not useful. Nevertheless, we find that in 40% of the 96,651 water points with payment information, users do not contribute financially to the system. Vandalism and theft emerged in a surprisingly large number of records, something which until recently has been relatively little remarked upon (Chowns 2014, 2015). Finally, the fact that we observed a much larger fraction of failure types being technical problems rather than management or institutional failures could indicate the initial planning, design and construction processes are working well overall but that points fail later for technical reasons, as would be expected. This would be consistent with the results in the literature on post-construction support, including Whittington *et al.* (2009).

We conclude with two thoughts on the ongoing global efforts to expand monitoring activities and share data publicly. First, the results from this study could be used to inform the development of a dropdown menu for non-functionality reasons, as opposed to an open response functionality status description in the WPDx database. Our time-intensive method of manually coding open response data is not sustainable if WPDx (or a similar enterprise) is successful and vast numbers of new uploads appear on the platform. Machine-learning approaches may be a productive avenue. In November 2016, WPDx teamed with HP Enterprise to develop a 'trainable' tool that can be used to extract patterns and data from the open-ended fields, and in fact used our draft codings to help train the algorithm (see https://www.waterpointdata.org/blog/2016/bringing-order-wide-open-fields).

More fundamentally, though, we feel our attempt to learn from the open-ended responses in the WPDx data points to the value of coordination in monitoring data programs and the inherent tradeoff in quantity of data gathered. Note too that one of the collaborative behaviors in the Sanitation and Water for All initiative is to 'strengthen and use country systems'; see http://sanitationandwaterforall.org/about/the-four-swa-collaborative-behaviours/. To gain first-mover advantage and become the dominant platform, non-profit organizations like WPDx and other data aggregators (Akvo Flow, mWater) have a difficult pitch to data contributors and funders. For contributors, the process of uploading and sharing data must be as painless as possible. This implies asking them to make few if any changes to their existing survey methodologies or questions asked. Better to rely on a 'core' set of questions that are relatively easy for governments and NGOs to provide. To funders, however, one of the promises of data aggregation is sector learning – specifically, how we can increase the sustained functionality of costly water system investments? The latter requires more than simply having geospatial information on water points along with their functionality status. Arriving at a common definition of 'functionality' and its levels will be a crucial step forward, and the most important purpose of monitoring is to help those responsible for services keep them operating and improving. However, one wonders whether in five years WASH sector professionals will still be looking at attractive, interactive maps of red, green and yellow dots but still be puzzling over why the red and yellow dots are not green. We may wish we had asked for more *structured* information on the 'secondary reasons' and 'underlying conditions of failure' shown in Bonsor *et al.* (2015)'s helpful Figure 1. What user fees are being collected? How regularly is the water committee meeting? How do users view the water services? Are spare parts currently on hand? Does the committee have external technical support? This implies moving beyond simply standardizing text fields (as in a dropdown) to standardizing a structured set of questions with built-in logic that can reliably be used to understand the most important proximate/technical reasons a point has failed as well as the underlying contributory reasons. Asking organizations to agree to a common set of survey metrics or even survey instruments is far more difficult, and this type of survey effort will require more enumerator time in the field for many organizations. However, the payoff to the sector from more in-depth coordinated data collection could very well be quite large.

ACKNOWLEDGEMENTS

We thank Brian Banks at WPDx, as well as Susan Davis, Rich Thorsten, Dale Whittington and two anonymous reviewers for helpful comments and discussions. We received no funding for this work and have no conflicts of interest to declare.

REFERENCES

Baumann, E. 2009 *May-day! May-Day! Our Handpumps are not Working!* RWSN Perspective No. 1, RWSN, St Gallen, Switzerland. Available from: www.rural-water-supply.net/en/resources-top/details/206.

Bonsor, H. C., Oates, N., Chilton, P. J., Carter, R. C., Casey, V., Macdonald, A. M., Alowo, R., Wilson, P., Tumutungire, M. & Bennie, M. 2015 *UPGro Catalyst Grant Report – A Hidden Crisis: Strengthening the Evidence Base on the Sustainability of Rural Groundwater Supplies – Results From a Pilot Study in Uganda*. BGS Open Research Report, OR/14/074 (94).

Carter, R. C. & Ross, I. 2016 Beyond 'functionality' of handpump-supplied rural water services in developing countries. *Waterlines* **35** (1), 94–110.

Carter, R. C., Harvey, E. & Casey, V. 2010 User financing of rural handpump services. In: *IRC Symposium Pumps, Pipes and Promises*. Available from: www.ircwash.org/sites/default/files/Carter-2010-User.pdf.

Chowns, E. 2014 *The Political Economy of Community Management: A Study of Factors Influencing Sustainability in Malawi's Rural Water Supply Sector*. Doctoral Dissertation, International Development Department, School of Government and Society, University of Birmingham.

Chowns, E. 2015 Is community management an efficient and effective model of public service delivery? Lessons from the water supply sector in Malawi. *Public Adm. Dev.* **35**, 263–276.

Danert, K., Armstrong, T., Adelike, D., Daffau, B., Quedraogo, I. & Kwei, C. 2010 *Code of Practice for Cost Effective Boreholes (June)*. Available from: www.rural-water-supply.net/en/resources/details/128.

Danert, K. T. 2013 *Experiences and Ideas From RWSN's Sustainable Groundwater Community 2013*. RWSN. Available from: www.rural-water-supply.net/en/resources/details/517.

Fisher, M. B., Shields, K., Chan, T. U., Christenson, E., Cronk, R. D., Leker, H., Samani, D., Apoya, P., Lutz, A. & Bartram, J. 2015 Understanding handpump sustainability: determinants of rural water source functionality in the Greater Afram Plains region of Ghana. *Water Resour. Res.* **51** (10), 8431–8449.

Foster, T. 2013 Predictors of sustainability for community-managed handpumps in sub-Saharan Africa: evidence from Liberia, Sierra Leone, and Uganda. *Environ. Sci. Technol.* **47** (21), 12037–12046.

Franceys, R. & Pezon, C. 2010 *Services are Forever: The Importance of Capital Maintenance in Ensuring Sustainable WASH Services*. (August) Available from: www.ircwash.org/resources/services-are-forever-importance-capital-maintenance-capmanex-ensuring-sustainable-wash.

Harvey, P. A., Skinner, B. H. & Reed, R. A. 2002 Sustaining handpumps in Africa: lessons from Zambia and Ghana. 28th WEDC Conference, Sustainable Environmental Sanitation and Water Services, Kolkata, India, pp. 101–104.

Harvey, P. A. 2004 Borehole sustainability in rural Africa: an analysis of routine field data. In: *Proceedings of the 30th WEDC Conference*. Loughborough University, UK.

Isham, J., Narayan, D. & Pritchett, L. 1995 Does participation improve performance? Establishing causality with subjective data. *World Bank Econ. Rev.* **9** (2), 175–200.

McPherson, H. J. & McGarry, M. G. 1987 User participation and implementation strategies in water and sanitation projects. *Int. J. Water Resour. Dev.* **3** (1), 23–30.

Rural Water Supply Network (RWSN) Executive Steering Committee 2010 *Myths of the Rural Water Supply Sector*. Available from: www.rural-water-supply.net/en/resources/details/226.

Rural Water Supply Network (RWSN) 2013 *Experiences and Ideas From RWSN's Sustainable Groundwater Community*. Available from: www.rural-water-supply.net/en/resources-top/details/517.

Rural Water Source Network (RWSN) 2014 *Water Point Mapping: Indicators, Pump Functionality, Accuracy of GPS, Using and Sharing Data* (June), pp. 1–9. Available from: www.rural-water-supply.net/fr/resources/details/588.

Schweitzer, R. W. & Mihelcic, J. R. 2012 Assessing sustainability of community management of rural water systems in the developing world. *J. Water Sanit. Hyg. Dev.* **2** (1), 20–33.

U-Dominic, C. M., Ezeabasili, A. C. C. & Okoro, B. U. 2014 Community-government partnership and sustainability of rural water programmes in Anambra State, Nigeria. *J. Environ. Earth Sci.* **4** (13), 1–13.

Whittington, D., Davis, J., Prokopy, L., Komives, K., Thorsten, R., Lukacs, H., Bakalian, A. & Wakeman, W. 2009 How well is demand-driven, community management model for rural water supply systems doing? Evidence from Bolivia, Peru and Ghana. *Water Policy* **11**, 696–718.

First received 27 January 2017; accepted in revised form 10 September 2017. Available online 19 October 2017