

Article

Wicked Water Points: The Quest for an Error Free National Water Point Database

Jeroen Verplanke * and Yola Georgiadou

Faculty for Geo-Information Science and Earth Observation (ITC), University of Twente, 7522 NB Enschede, The Netherlands; p.y.georgiadou@utwente.nl

* Correspondence: j.j.verplanke@utwente.nl; Tel.: +31-53-4874-532

Received: 5 June 2017; Accepted: 3 August 2017; Published: 8 August 2017

Abstract: The Water Sector Development Programme (WSDP) of Tanzania aims to improve the performance of the water sector in general and rural water supply (RWS) in particular. During the first phase of the WSDP (2007 to 2014), implementing agencies developed information systems for attaining management efficiencies. One of these systems, the Water Point Mapping System (WPMS), has now been completed, and the database is openly available to the public, as part of the country's commitment to the Open Government Partnership (OGP) initiative. The Tanzanian WPMS project was the first attempt to map "wall-to-wall" all rural public water points in an African nation. The complexity of the endeavor led to suboptimal results in the quality of the WPMS database, the baseline of the WPMS. The WPMS database was a means for the future monitoring of all rural water points, but its construction has become an end in itself. We trace the challenges of water point mapping in Tanzania and describe how the WPMS database was initially populated and to what effect. The paper conceptualizes errors found in the WPMS database as material, observational, conceptual and discursive, and characterizes them in terms of type, suspected origin and mitigation options. The discussion focuses on the consequences of open data scrutiny for the integrity of the WPMS database and the implications for monitoring wicked water point data.

Keywords: rural water supply; water point mapping; Tanzania

1. The History of Water Point Mapping in Tanzania

Tanzania has a rich history in rural water supply. Jiménez and Pérez-Foguet [1] describe how, before independence, rural water schemes were implemented mainly by national government, while local government was responsible for operation and maintenance through water fees and taxes. Shortly after independence, the new government of Tanzania decided that all costs should be borne by the government and public water should be free. During the 1980s, a new policy made water users responsible for operation and maintenance of water schemes, and donors contributed significant funding [1]. Over the course of the 1990s, new targets were set to achieve rural water supply service to within 400 m of all households by 2002.

The Government of Tanzania and international Non-Governmental Organizations (NGOs) have faced many challenges in the devolution of responsibilities for rural water supply to local actors. Jiménez and Pérez-Foguet [1] observe that the practice of devolving responsibility for rural water supply without simultaneously devolving the necessary financial resources and instituting coherent policies to support this, has persisted since independence, despite the warnings of academic literature [2–4].

In 2009, the Ministry of Water and Irrigation (MoWI) released the Water and Sanitation Act [5], which promulgated a "demand-response approach," whereby "the central government plays the role of coordinator and facilitator in the water sector, and the district level holds the main responsibilities for implementation" [1]. This approach to service delivery depends on communities to demand, own,

and maintain their water services and participate in their design, as well as to be responsible for operation and maintenance costs.

Securing the resources for operation and maintenance has been very difficult for most rural communities and nearly impossible without external funding. The sustainability of water services is further jeopardized by the low level of professionalism in the management of services [6], the difficult relationship between water users and elected representatives and the limited role that local and district governments play in the monitoring of water point functionality and the provision of technical support [1].

The purpose of the Water Point Mapping System survey [7] was to collect for the first time ever a baseline of accurate, reliable and up to date information on all water points (WPs) in rural Tanzania. Like all baselines, this particular one was supposed to underpin not only the monitoring of all functional and non-functional public rural WPs at any future time, but also to improve decision-making and allocation of resources, leading to improved water supply services in rural areas. Several databases preceded the WPMS. The German Development Agency-GIZ set up a Rural Water Supply (RWS) database in 2001. The purpose was to record information on existing water schemes in rural areas. The system worked well and tracked information at the district and ward levels and was used as input for national policy reports [8]. The data was updated through paper forms that were filled out in the field and manually entered into the database. The RWS database contained 2765 schemes when it was last updated in 2007. The World Bank also set up the Maji Management Information System (MIS) in 2004, which was used until 2008. The Maji MIS covered only 14 districts and was essentially a project management tool comprising the procurement, construction and financing of rural water schemes. Neither the Maji MIS nor the RWS systems are linked to the current WPMS.

Water point mapping (WPM) was initiated in Tanzania by Water Aid in 2004 [9,10] to scale this NGO's previous positive experiences in Malawi. Upon seeing the outputs of the water point mapping exercise, the Permanent Secretary institutionalized the process within the MoWI. Water points in fifty-five out of 132 Tanzanian rural districts were mapped between 2005 and 2009 using broadly the Water Point Mapping (WPM) methodology championed by Water Aid, and adopted by other actors in the international water sector (SNV, Plan International, Concern Worldwide).

The outcomes of these WPM efforts were fed into discussions at national sector review meetings [11]. By 2008, stakeholders in collaboration with the MoWI had successfully legitimized WPM as a useful monitoring tool and revealed a 43% functionality rate among mapped water points in rural areas [12]. The results of four case studies [13] showed that the main constraints were the lack of updating mechanisms, lack of use as well as lack of integration of the system with the other systems in the decision-making and planning process [1]. SNV then carried out a Validation and Inquiry Process (VIP) [14] to investigate why so many water facilities were not functioning.

From 2010 to 2013, the MoWI commissioned a consultant to carry out a Water Point Mapping (WPM) project in all districts in Tanzania, to monitor the functionality performance of rural water supply schemes and water points [15]. The purpose was to build on existing experiences and benefits obtained from the Water Aid experience with the view of improving decision making and allocation of resources towards improvement of water supply services in rural areas. According to the specifications of the WPM project, the consultant had to: (1) locate each rural WP in Tanzania by Global Positioning System (GPS); (2) take pictures of each WP; (3) collect data on the functionality, management, specifications and water quality and quantity of each WP. The WPM project also included a web-based GIS system to produce and make publicly accessible maps and data relating to WP functionality and coverage. Further, the project should facilitate an increase of the capacity of the MoWI and local government staff to use and update the WPM database and other stakeholders in the country to understand the status of rural water supply services in terms of coverage and functionality [12].

2. The Need for Water Point Mapping

The Sector Programme for Rural Water Supply in Tanzania (2006) set goals for the percentage of the population in rural areas with sustainable and equitable access to safe water. The first goal of

65% coverage was to be achieved in 2010 [16], and should subsequently grow to at least 74% by mid-2015 to comply with the Millennium Development Goals (MDGs) for access to water.

In 2010, the MoWI set certain annual milestones such as “countrywide quarterly functionality monitoring of all water points in Tanzania” [15] in order to create a baseline that could be used for results-based reporting of outcomes. Until then, the monitoring of the MoWI depended on routine output data to calculate service coverage. In this data, water service coverage was based on the number of constructed water points per 1000 inhabitants to calculate an assumed number of persons served. However, the data did not record whether the water installations were functioning. The MoWI and the National Bureau of Statistics noticed this flaw and decided to adopt an outcome based monitoring approach: “actual access rates are likely to be less (and possibly much less) than those reported using routine data. The reason for this discrepancy is clear: routine data does not record functionality and assumes that investments do not fail. Outcome (access) surveys do record situations where water points (or entire schemes) have failed for technical, financial, management or any combination of shortcomings. Without a reliable baseline that takes into account functionality and (more importantly) a means to keep this updated, it is impossible to track the net progress in expanding rural water supply service coverage or, more importantly, to determine actual access rates” [15].

Non-conflicting data on the number of water points available and the rural population served fueled the need for Water Point Mapping [17]. Originally conceived as a planning and budgeting tool to encourage the transparent and evidence based allocation of resources, Water Point Mapping was later also seen as an excellent tool for communities and local leaders to visualize a rural water scheme and its challenges. It was therefore envisioned that the baseline data could be updated by Community Water Supply Organisation (COWSO) representatives using their mobile phones, similarly to the Human Sensor Web system tested on Zanzibar [15,18].

The contract to collect baseline information was supposed to be completed by December 2011. By then, a functionality tool would be operational in all 132 LGAs. By the middle of 2012, the project had not yet finished due to internal delays in the disbursement of funding. By that time however, producing accurate and up-to-date data on rural water supply infrastructure was seen “as one of the most urgent challenges facing the sector” [19]. The MoWI had established that it could enable government and other stakeholders (if authorized to do so) to monitor and analyze functionality and other aspects of all water points in real time, via a web based interface and even establish “the status and reasons for non-functioning water supply and identify rehabilitation requirements” [19].

At the same time [19], the Government of Tanzania committed to the Open Government Partnership (OGP), a global initiative aiming to promote transparency and citizen empowerment, to fight corruption and to encourage the use of new technologies to improve governance, e.g. platforms to engage citizens through the internet and mobile phones to monitor and report water point functionality to government. The OGP demanded that the disaggregated data from the WPM become available online (in machine-readable format) in order for “local government authorities to use data to plan for new investments and NGOs to use these data for planning their own investments [and] researchers to increase understanding of sustainability and equity issues for the water sector in the country” [19]. Thus, WPM was framed as a tool to produce accurate and timely data that could be disseminated through user-friendly maps and reports.

The consultant was eventually able to fulfill his contract mid-way through 2013 [17]. By then, a total of 75,777 water points had been mapped, of which 46,697 water points were found functional (62%) and 29,080 were found not functional (38%), as reported on the MoWI website. At the same time, only 200 COWSOs had been established in the 132 Local Government Authorities—a tiny fraction of the total number of COWSOs required nation-wide to update the water point data in the future.

3. How Are Water Points Wicked?

“Wicked problems” is a popular concept in policy and information sciences. Several characteristics of wicked problems, defined by Kunz, Rittel and Webber [20,21], are relevant for rural water services. First, the framing of policy problems is not universal—“public water service” in

France is roughly similar but discretely different from that in Tanzania. Second, it is difficult to achieve consensus regarding the solution of wicked problems—some may claim that constructing more public water points will improve water service, while others counter that genuine decentralization of public services is the solution. Third, solutions to wicked problems can only be subjectively better or worse, not objectively true or false—water service provision should be incrementally improved, rather than solved at one stroke forever. Fourth, many cause-effect stories can be advanced for a wicked problem, depending on the individual perspective of the stakeholder. As Rittel and Webber put it: “the information needed to understand the problem depends on one’s idea for solving it” [21], where access to water can be improved from a health care or sanitation perspective but also from the point of view of a basic human right. Fifth, every wicked problem is a symptom of another problem—reduced school attendance for girls in many African countries is connected to the time required to collect water [22,23]. Sixth, proposing a solution to a wicked problem frequently prevents incremental design because most interventions change the original problem—introducing payment for (improved sources of) water in Tanzania to finance the maintenance of water points has caused many people to resort to the use of unimproved (free) water sources [24].

Aligning public water services with wicked problems is done in a different way by Rottenburg [25] in what he calls the “technical game”. Rottenburg discusses in his parable of development aid the irresolvable internal contradiction in international development cooperation. He describes the “accountable, predictable and obviously conditional transfer of resources from the North to the South versus the facilitation of sustainable and self-determined development of target countries”, which requires vast amounts of quantitative data to be supplied in order to provide proof of progress and achievement. This is reflective of the mentioned characteristics of wicked problems. To resolve this contradiction, development partners from the North and the South play a “technical game” that brackets the local social and cultural frames of reference. Development partners no longer focus on the wicked problem of improved rural water supply but only on the (seemingly) tame problem of mapping the distribution of that rural water supply or the mere “production” of water points in villages without considering the availability of other water sources.

As this research is part of a larger investigation funded by the Dutch Science Foundation, we also use the framework of Pritchett and Woolcock [26] and the World Bank [27] as adapted by Nganyanyuka et al. [24], which distinguishes between discretionary and transaction-intensive elements in key services to citizens. Transaction-intensive elements, like mapping or monitoring all rural water points in Tanzania, require a large number of transactions, involving face-to-face contacts between district officials, village water technicians, COWSO members, and citizens—for example, a water technician detecting a broken water point and reporting the breakdown to the COWSO secretary. Discretionary elements involve decisions based on information that “is important but inherently imperfectly specified and incomplete, and entails extensive professional or informal context-specific knowledge” [26] (p. 194). Collecting and digitizing data about transaction-intensive elements of water services is relatively easy, while collecting data about discretionary elements is fraught with insuperable difficulties [25]. It is precisely the discretionary nature of water point mapping that renders the water points “wicked” and their mapping a “wicked problem.”

4. The Water Point Mapping Data

We analysed the various attributes of WPs captured during water point mapping (WPM) in Tanzania from 2010 to 2013 and recorded in the online database of rural water points dated 25 April 2013 (APR.2013). In mid-February 2014, a new version of the WPM data was published on the Ministry of Water’s website and denoted FEB.2014. Both versions were officially available on the government website on 21 May 2014. The version FEB.2014 is different from the previous version (APR.2013) in several ways. The FEB.2014 version was organized in spreadsheets, one spreadsheet per Tanzanian region, and contained close to 68,000 water points—about 7,500 fewer water points compared to APR.2013. In the new version, however, one region with 6293 water points was excluded. The total number of mapped water points in Tanzania including this region therefore

amounted to 74,250; 1500 less than the original data set. The FEB.2014 version had, however, been considerably cleansed of duplicate records. The APR.2013 version featured over a thousand duplicate records of which only around 100 duplicate records (coordinates) remained in the FEB.2014 version. Database cleansing has most likely led to the reduction of the total number of water points. The new data also contained information on the geographic coordinates. The previous version lacked metadata regarding the map datum used to collect the GPS coordinates of the water points. The table headers in the new spreadsheets indicate that the data were collected using the Arc.1960 map datum. As geographic overlays of water point (WP) information and administrative boundaries show some strange overlaps (water points positioned in neighboring districts), the map datum information can be used to correct these map errors. In 2015, several updates of the WPM data were made. The data, however, became less accessible to the general public due to the transfer of the database under the OGP as the raw data tables were no longer downloadable. Only pdf files and data per administrative region could still be visualized on an interactive website. As most evaluations of the Tanzanian Rural Water Supply have been based on the WPM data, initially made available (APR.2013 version), this paper is basing its argumentation on this dataset as well.

4.1. Errors in WPM Data

To analyse the errors in the WPM data and to assess the issues arising from these errors, we examined the attributes of the WPM data (spreadsheet columns), and catalogued and classified the anomalies in the data according to error types. If we see the WPM survey as an experiment, a common way to look at errors is to classify them as systematic, random and gross errors [28–32]. Systematic errors, e.g., due to wrong calibration of instruments (e.g., settings in a GPS), can be eliminated through recalibration; random errors may be estimated statistically [31,32]; blunders or gross errors, made when values are incorrectly selected or marked [29]. Besides deletion, no other solution exists for the correction of gross errors. Deletion of gross errors therefore leads to loss of data and the only way to get the data cleaned is to repeat their collection in the field.

Nevertheless, errors encountered in the WPM data, cannot be classified in this way since a mapping project is unlike a scientific experiment. The variables measured by an instrument lack metadata regarding the applied method and procedure and make it difficult to assess whether an error is systematic. Secondly, the effect of random errors on the data cannot be quantified easily because the individual water points are unique, independent features in the landscape and their attributes are uncorrelated. Of course, there will be some form of correlation between some of the recorded attributes, like the water quality of water points connected to the same aquifer. However, these errors mostly occur in unrelated elements like the order in which water points are named, the number of people using a water point, the quantity of water measured, the type of pump technology, and the level of point improvement. Even the donor agency (another WP attribute) cannot be deduced from neighboring water points. Many water points are either donated by individuals or by multiple donors active in a village consecutively or together for several years. Thirdly, many errors seem to arise during data entry or from ambiguity. Spelling errors and contradicting columns suggest the existence of significant gross errors. Much of the analysis of the data and a consequent judgement on the integrity and usability of the data therefore comes down to studying these gross errors.

Allchin [33] provides a classification of errors that is dependent on observational benchmarks derived from both fact and theory as well as local cultural context. As this classification also resonates with our theoretical framework of transactions and discretion [25,26], we have chosen to adopt it. It creates a contextual spectrum of errors that can be classified into material, observational, conceptual and discursive errors:

- **Material errors** can be caused by improper procedures (violation of protocol or poor technical skills) and involve “aspects of getting the phenomenon right” [33]. In WPM, this encompasses (a) the filling of the data entry form, (b) the use of GPS for water point location, (c) the use of water quality testing kits and (d) data processing.
- **Observational errors** occur when insufficient controls exist to establish domain observation or an incomplete theory of observation exists, reflected by the poor choice of instruments or field

methods. Observers can also exhibit a perceptual bias that is either “theory-laden” [33] or a problem of framing of the phenomenon. In WPM, this error type is reflected in (a) the choice of field equipment, (b) experience of the Water Point Collector (WPC) and (c) the management of the WPC team.

- **Conceptual errors** are commonly miss-specified assumptions or boundary conditions. They involve theoretical interpretations common in philosophy. The possible cognitive bias due to theoretical entrenchment is important. In WPM, this error type is due to (a) the rigidity of the data entry form, (b) changes over time in the WPM approach and (c) the framing of rural water service problems by stakeholders at different levels of the Water Point Mapping System.
- **Discursive errors** can originate from communication failures (incomplete reporting, translation) or mistaken judgments of credibility but also from unchecked sociocultural cognitive biases and public misconceptions. In WPM, these problems arise with (a) the intelligence sources for the different WPM attributes, (b) misunderstanding of WPM concepts by local water users, (c) misinterpretation of local knowledge by the WPC and finally (d) fraudulent data manipulation.

This classification, although conceived to analyze scientific results, enables a narrower interpretation of the WPM survey results. Particularly, the assumption that the benchmarks of observation have a local cultural aspect is important when scrutinizing data from a field experiment such as WPM. This is linked to what Allchin [33] calls “second-order errors”. It involves the ability of local (scientific) institutions to warrant claims and produce knowledge effectively with the ability for error remediation.

In Appendix A, 36 of the collected attributes in the WPM data are displayed in relation to the intelligence sources and their potential problem manifestations (Table 1). We identified potential problems with 26 of the 36 attributes. These problems were categorized according to the error typology we adapted from Allchin [33]. We defined 14 different root causes within these four error types that corresponded with suspected errors. Many of these cases occur, however, in combinations of different error types derived from the manifestation of the errors:

1. material errors
 - a. filling of the data entry form (in 19 attributes)
 - b. use of GPS (in one attribute (for GPS, there are many attributes (>20) recorded with the WPM data, as these are all automatically recorded they have been grouped as one. In the open WPM dataset lat-lon coordinates (two attributes) and GPS height (one attribute) are presented without their error values, making it impossible to calculate systematic or random errors in the location measurements. The quality of GPS data therefore in this analysis depends only on the proper use of the device (1b), the choice of equipment (2a) and experience of the operator (2b))
 - c. use of water quality testing kits (in one attribute)
 - d. data processing (three attributes)
2. observational errors
 - a. choice of field equipment (four attributes)
 - b. experience of the Water Point Collector (WPC) (seven attributes)
 - c. management of the WPC team (consistency and training) (two attributes)
3. conceptual errors
 - a. rigidity of the data entry form (nine attributes)
 - b. changes over time in the WPM approach (three attributes)
 - c. framing of rural water service problems by stakeholders at different levels of the Water Point Mapping System (five attributes)
4. discursive errors
 - a. intelligence sources for the different WPM attributes (eight attributes)

- b. misunderstanding of WPM concepts by local water users (three attributes)
- c. misinterpretation of local knowledge by the WPC (seven attributes)
- d. data manipulation (two attributes)

Table 1. Broadly suspected causes of error in Water Point Mapping data categorized according to error typology adapted from Allchin [33].

Error Type (Allchin, 2001)	Suspected Origin	Manifestation	Mitigation Options
material error	human computer interaction (office)	duplication of records syntax error	Can be solved relatively easily
observational error	human computer interaction (field)	missing data, duplication, ambiguity through miss typing or touch/tap error in preformatted fields	Difficult to solve. Hard to establish whether the chosen option was correct or not
conceptual error	human-human interaction (office)	changes over time, procedural rigidity, definition of functionality	Hard to solve after data collection has commenced. Requires database adaptation or changes in data collection strategy
discursive conceptual	human-human interaction (field)	subjective information. information not matching between columns (functional and breakdown) = conflicting information	Difficult to solve. Hard to establish whether the source was authoritative or not
material and observational error	pure human error (e.g., forgetfulness)	regular, repetitive mistakes in numbering or misspelling (field)	Can be solved, but may require considerable resources
discursive error	malicious intent	forgery of records (field and office)	Difficult to solve. Unless geographical coordinates are forged badly or unnatural patterns are visible in the data

When translating the manifestations of error into Allchin's error types, it becomes clear that very few of the errors pertaining to specific attributes can be uniquely classified. In most attributes, errors are of mixed types if not all four types. If only one type would be found, the root cause would likely be either a person, a tool, a method or a line of reasoning. In reality, however, these errors are due to a cascade of causes starting with a wrong line of reasoning that leads to a poor choice of method, which, in the end, leads to ambiguity in the collected information, as the WPC is unable to fit the observations into the possible options provided in the data collection form. The ambiguity in the resulting information could be the starting point for new or poorer lines of reasoning resulting in a vicious circle. Another example could be a faulty tool that leads an inexperienced WPC to believe he is operating within margins.

Most of the errors identified in the Tanzanian WPMS are neither intrinsic to Water Point Mapping nor to the Tanzanian cultural context. The material and observational errors may occur in any large spatial data collection survey, particularly those conducted at a national scale. The discursive errors may be attributed to a wider developing country setting where local capacity is insufficient to provide the required support or information. Only the conceptual errors originating from the framing of water service issues have the specific local context as a root cause. The rigidity of the data collection form appears to be the result of the many requirements set by the MoWI for the WPMS, which left little space for the consultant to maneuver. The Tanzanian framing of what constitutes a functional water point (see Section 4.2.6) and the rearrangement of local government authorities during the duration of data collection are causes of error that can only be attributed to the specific national context.

4.2. Manifestation of Errors

As shown in Table 1, errors in the data are due to a number of reasons. These are usually combinations of the four error types. The causes and effects of these errors can be inferred by discussing the errors based on their manifestation: changes that have happened over time, syntax error in the input of the data, missing data and ambiguous values, subjective observations by field

operators, duplicate records created at different occasions during data handling and the definition of functional water points.

4.2.1. Changes over Time

WPM data was collected over a three-year period from 2011 to 2013. Around 33,300 points were collected in 2011, 5200 points in 2012 and the remaining 27,000 in 2013. The data also includes 37 water points recorded as early as 2004. The first WPM exercise (by SNV, Water Aid, ISF and Concern) contained around 24,500 water points recorded between 2002 and 2009. These points were revisited and updated in the period 2011–2013. This implies that the information for almost half of the water points (33,300 out of 68,000) was already three years old at the moment of publication.

The fact that the information was collected over a period of three years does not seem to have affected the consistency of the approach. The data collection form used for WPM was adapted from the form used in the first exercise led by SNV [14] by adding a few categories to the form, and removing none (Box 1). This could have allowed for seamless updating of the pre-recorded water points.

Serious challenges for the consistency of the data, however, occurred during the survey time span (2011–2013). These were caused by the renaming and numbering of local government authorities (LGA) in 2012 [34] and the related merging of administrative wards within LGAs and resulted in the disconnection of whole villages in the database. These errors in the data can only be removed by updating the water point records taking the unique identifiers of the water points as a starting point. Checking these against the actual ward and LGA names could correct this problem. These unique identifiers, the water point codes, are, however, one of the major challenges in the data as explained in the next section.

Box 1. Changes made to the Water Point Mapping data collection form between 2010 and 2013.

Added attributes in the WPM data collection field form:

- Local Government Authority name, Village population, Village photo, WPT code, Population served by WP, Catchment name, Existence of Water permit, Year of construction.
- Existing Attributes with added options are:
- Source type; added two options for “rain water harvesting, roof or ground”, previously only rain water harvesting
- Extraction system; added “SWN 81” and “India Mark III”
- Status; added four classes of functionality: “Functional needing repair, not functional >6 m, not functional <6 m, not functional <3 m”
- Hardware problem; added “Hand-Pump broken” and “on rehabilitation”. The option “other reasons for not functional” was renamed to “reason for not functioning”
- Water quantity; added “others”
- Water Payments; added “amount Tsh”

4.2.2. Syntax Error

Each water point (WP) has a unique identifier (WPCODE). The syntax of the WPCODE is created by consecutively combining the numbers of Region-District-LGA-WARD-VILLAGE (11 digits) + a WP number, e.g., “01020030405WP001”, totaling 16 characters. The added “WP001” is created by the WPC: the first WP encountered in a village receives #001 [12]. The data shows that the 16-digit syntax is only observed for about 50% of the water points. In wards with less than 100 WP (which is most of them), the first WP is numbered WP01 creating a code with 15 characters only. While this error can be easily mitigated in the database, other “zeroes” in the identifier were also omitted (e.g., when a LGA, Ward or Village was numbered “012”). As a consequence, the location (up to village level) of about 50% of the water points cannot be logically inferred anymore from this unique identifier. Once the WPC has formulated the WP code, it was physically recorded on the WP with either paint or a tag. The data does not allow the analyst to assess whether the syntax errors are replicated on the water point tags. Only a revisit in the field of all the points with wrongly formatted codes could

provide a conclusive answer. Furthermore, the data contained many duplicate records. In most attributes, it seems that these duplications occurred when the WPC moved to another village or ward and restarted numbering WP001 without changing the ward or village numbers accordingly.

4.2.3. Missing Data and Ambiguous Values

The data contains several fields that raise questions. It is not always clear whether a “zero” (0) actually represents a value of zero or whether it reflects lack of data. The data for instance contains information on available PRIVATE CONNECTIONS, but only 952 records are available. Without metadata, it cannot be assessed whether the missing value “0” represents the fact that no private connections exist or whether no data was available. As many rural areas are also serviced by urban water supply schemes, it is possible that many more private connections exist nationwide than those recorded.

Further, many data fields are empty. Out of 25,209 non-functional water points, only 16,546 points provide information on BREAKDOWN YEAR. Collecting this attribute obviously requires local knowledge. Apparently, either nobody knew when those water points had broken down or the WPC was unable to consult anyone who knew. Additionally, the data features several records where the breakdown year precedes the year of construction. In addition, 1010 WP in Mwanza are missing GPS coordinates. If some of these WP also have wrongly formatted ID Codes, they can be presumed “lost” and cannot be retrieved for updating.

Other data is completely missing. The FEB.2014 dataset no longer contains the information on the recording date of the WP, making it harder to assess which WP should be prioritized in the updating. One piece of information that was originally missing in the data (APR.2013) was the metadata on the map datum used for the GPS recordings. Mapping the WP data in a GIS without datum information lead to challenges caused several WPs to be located in the wrong administrative unit or even within water bodies (e.g., lakes) when using the default map datum settings of the GPS devices. The FEB.2014 dataset fortunately provided this information in the column headings, indicating that the map datum used was Arc.1960.

Ambiguity is evident in the information provided with FUNDER, INSTALLER and YEAR_OF_CO. These attitudes are commonly recorded (painted) on the WP base/slab upon completion of installation and are copied by the WPC on the spot [12]. Very often, the information had worn off or was damaged and had to be provided by a member of the COWSO. Whether copied from the WP or recorded from the COWSO member, the WPC entered this information manually. Since many WPC were involved in the water point mapping, a multitude of spellings of names under FUNDER and INSTALLER were used. For instance, WP funded by the German Government may have been recorded as either: “Germany, german, ger or G”. “G”, however, may represent the Government of Tanzania. Such different spellings are common in the database for every donor, rendering much of this information useless.

In some attributes, the information leads to suspicion about conflicting information between the observations made by the WPC and the information provided by his local informants. The attribute HARDWARE_P (“Main hardware problem”) was recorded based on predefined reasons for WP breakdown. The next attribute on the list “REASON_WPT” was originally defined in the 2010 SNV Data Collection Form as “other reason WP not functional”. During the WPM data collection, however, the WPCs interpreted this as either “the reason for the recorded hardware problem” or, if the WP was judged as functional, but other problems were evident, as “additional reasons” [12]. There also was a final attribute to record “General comments”. These were provided by the WP collectors themselves and are general statements about the WP. These comments (Figure 1) sometimes give valuable insight into the situation on the ground, indicating whether management of the WP was poor or describing the “actual” problem, from the point of view of the WPC, if the available options of the data collection form did not include it.

- Functional;
- Functional needing repair;
- Non-functional <3 months;
- Non-functional >3 months;
- Non-functional <6 months;
- Non-functional >6 months;
- Non-Functional.

STATUS 2 is an aggregate done by the consultant during data processing, when he merges these seven classes into only two classes:

- Functional (including Functional and Functional Needing Repair) and,
- Non-functional (including all other classes).

Based on the number of functional water points, the rural water supply coverage can be calculated.

The Second National Strategy for Growth and Reduction of Poverty (MKUKUTA II) required the national government to increase the access to clean and safe water supply from 58.7% in 2009 to 67% in June 2015 [16]. The first Rural Water Supply and Sanitation Programme (2006) established much higher targets for the population in rural areas with sustainable and equitable access to safe water: 65% by 2010 (MKUKUTA), at least 74% by mid-2015 (MDGs), and 90% by 2025 [1]. Table 2 shows that the improvements over the last years have, however, been small and coverage seems to decline since 2009. In 2013, however, the Baseline for Rural Water Service Coverage (RWSC) was downgraded from 57% to 40% [35], which leaves an even more considerable gap to cover in the remaining years.

Table 2. No. of people with access to water supply service in rural areas between 2009–2012.

Year	District Population (Dp)	No. of People with Access to Water Supply in Rural Areas (Ps)	% of Rural Water Service Coverage $W_c = P_s/D_p \times 100$
2005	30,995,135	16,308,651	53.74
2006	33,767,106	18,798,723	55.7
2007	38,337,892	21,675,360	57.10
2008	39,105,062	22,790,460	58.3
2009	33,536,205	19,685,659	58.70
2010	35,569,876	20,545,945	57.76
2011	36,474,939	20,634,227	56.57
2012	39,413,223	22,443,769	57

The reduction in the baseline of 2012 from 57% to 40% can be partly explained based on the variability of the number of improved water supply infrastructures that are currently functional. The variation is caused by the multiple interpretations of Functional Water Supply Coverage. For starters, the formula for calculating functional coverage in Tanzania assumes one functional water point for every 250 people. Hence, four functional water points per 1000 people results in 100% water supply coverage. The problem with this formula is to agree on the definition of “functional”, a problem shared by all specific indicator definitions. Proper definitions should involve issues of affordability, quality, reliability and non-discrimination—“exactly identifying what should be measured remains challenging” [36]. Finally, the calculation of the Rural Water Service Coverage depends on whether to include only fully functional water points or also those needing repairs.

5. Conclusions

The Open Data policy of Tanzania created space and opportunity for data analysts to scrutinize the database of the water point mapping system and raised doubts about the integrity of data collected in projects funded by development partners. Data previously (before the Open Government Initiative) considered undisputed can no longer be taken for granted as input for achieving the new Sustainable Development Goal for access to water.

The material, observational and discursive errors identified in the WPMS database are not intrinsic to either Water Point Mapping or the Tanzanian context. These errors may occur in any data collection project in the global South, particularly those projects conducted at a national scale or with a mind-set from the global North expecting an enabling environment for any sort of enumeration. Countries that lack the resources or the benefit of decades and centuries of dedicated mapping and monitoring at a national scale cannot easily enumerate infrastructural elements in remote areas. On the other hand, the root cause of conceptual errors is the specific local context. Notably, a consensual definition of water point functionality in a database is fundamentally contextual and cannot be addressed as a universal challenge to Water Point Mapping. Neglecting the local context—in other words, neglecting the discretionary nature of water point mapping—rendered the water points “wicked” and their mapping a “wicked problem.”

The rationale for neglecting the local context is well established in the literature of development aid, most recently in Rottenburg’s parable of development aid [25]. International development cooperation requires an accountable, predictable and conditional transfer of resources from the global North to the South [25] as well as vast amounts of quantitative data in order to provide proof of progress and achievement in the water sector. To resolve the paradox of accountability to the North and project effectiveness in the South, development partners are inclined to transform the wicked problem of improving rural water supply to the more manageable problem of mapping the geographical distribution and attributes of water points. However, wickedness can strike back, as we have seen, no matter how diligently it is avoided. Thus, the representation of water points in the database turned out to be “wicked”—water points were duplicated, obfuscated, disappearing and often useless. As a result, the database construction became an end in itself instead of a means for the future monitoring of all rural water points and rural water supply.

Surveying, mapping and recording the 75,777 rural water points in Tanzania was a formidable task for the consultant. It required employing and training of many data collectors, dealing with changing policies, international politics and local administrative shifts, as well as interacting with a wide range of stakeholders—multiple levels of government, (international) NGOs, donor agencies, data collectors, COWSOs and local water users. Ambiguity, issues of authority and subjectivity, changes over time, procedural changes versus rigidity, and multiple definitions of the same concept (e.g., functionality) all fit the characteristics of wicked, or messy problems, as Horn and Weber [37] prefer to call them.

The question remains as to whether our findings are applicable in other policy domains. For instance, while the health and sanitation sector is similar to the water sector in terms of data collection requirements, the organization of data collection is inherently different. Health and sanitation professionals are required to conduct daily monitoring and reporting. Their data collection skills have been honed over years of continued and constantly improved monitoring protocols and alleviate many of the hurdles shown here for water point mapping. It is precisely for this purpose that the SEMA research project [24] focused on using reporters who are embedded in the local context of rural water supply to do the monthly monitoring of functionality.

A further parallel may be drawn to efforts like Humanitarian Open Street Map (OSM). Many individuals, some with little training, are tasked to add features to maps available online. These maps are becoming increasingly important in the global South, as they often contain more detailed information than maps of national mapping agencies. However, the types of errors discussed in this paper are usually detected or avoided before the information is published online. OSM data collection undergoes a strict verification of mostly transaction-intensive data. All map entries are verified by the mapping community and follow globally applicable and well-documented guidelines of data entry and verification. Although efforts are under way to agree on a global definition of water point functionality [38], a verification system comparable to OSM is not likely to be available any time soon.

Acknowledgments: This study analysed the Water Point Mapping database, made available online by the Government of Tanzania, as well as interviews with and reports written by national and international consultants involved in the Tanzanian Water Sanitation and Health sector. Particular thanks go to Mr Singolile Mwamwaja who gave us an eye-witness report of practices of data collection from the Water Point Mapping in

Tanzania in his capacity as a water point collector. We also acknowledge the insights into global experiences of water point mapping and the contextual explanations of these experiences provided by the consultants, technicians and academics interacting through the Rural Water Supply Network (RWSN). Research for this paper was part of an integrated research project, titled “Sensors, Empowerment and Accountability in Tanzania (SEMA)” and funded by the Netherlands Organisation for Scientific Research (NWO–WOTRO).

Author Contributions: Jeroen Verplanke analysed the data; Jeroen Verplanke and Yola Georgiadou wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Field data (attributes) collected during water point mapping in Tanzania from 2010 to 2012.

Attribute in the Data Collection Form		Intelligence or the Attribute Provided by	(Potential) Problem Manifestation	Error Type (Allchin 2001)
DATE_OF_RECORDING_	Recording date of water point	Water point Collector (WPC), WPM Device	Discrepancies between DATE_OF_RECORDING and GPS_DATE (see 4.2.3)	Material (1d) Discursive (4d)
LGANAME	Local Government Authority name	Village executive officer (VXO)	Problems arise when comparing different datasets as in 2012 many LGAs were renamed. (see 4.2.1)	Conceptual (3a)
WARD	ward name	VXO	Problems arise when comparing different datasets as in 2012 many Wards were merged (removed) or split (new) (see 4.2.1)	Conceptual (3a)
VILLAGE	village name	VXO	No problems	
VILLAGEPOP	Village population	VXO	Difficult to check consistency (use of 1, 0 or missing value) as only one value entered for all WP in village (look-up value in other row). Not all villages have pop value. Unclear if missing values represent no data (see 4.2.3)	Material (1a) Observational (2a) Conceptual (3a)
VILLREGNO	Village registration number	VXO	Many inconsistencies in numbering format. Propagates into Water Point (WPT) code (see 4.2.2)	Material (1a) Conceptual (3b)
VILLAGEPHOTOI	Village photo ID number	Water Point Collector (WPC),	Copied from digital camera. Many inconsistencies and missing values (see 4.2.3)	Material (1a) Observational (2c)
NO_PRIVCON	Number of private water connections in the village	District Water Engineer (DWE)	Unclear if missing values represent no data or no connections (see 4.2.3)	Material (1a) Conceptual (3a) Discursive (4a)
SUBVILLAGE	Name of sub-village	VXO, Village Water Committee (VWC)	No problems	
WPTNAME	Name of water point	VWC	No problems	
WPTCODE	Water point code	Ministry of Water (MoW), National Bureau of Statistics (NBS), WPC	Many inconsistencies in numbering format and duplications. (see 4.2.2)	Material (1a) Material (1d) Conceptual (3b)
POPSERVED	Population served by WP	VWC	Difficult to check consistency as value of 1 means either NO DATA AVAILABLE or POPSERVED is recorded with other WP in Ward to prevent double counting (see 4.2.3)	Material (1a) Observational (2a) Conceptual (3a)
WPTPHOTOID	WP photo ID number	WPC	Copied from digital camera while at WP. Many inconsistencies and missing values (see 4.2.3)	Material (1a) Observational (2c)

Table A1. Cont.

SCHEMENAME	Name of water scheme	DWE, VWC	Many missing values (see 4.2.3)	Material (1a) Discursive (4a)
WATERPERMI	Water Permit issued for scheme	Catchment authority	By definition (“yes”) is given	Conceptual (3a)
CATCHMENT	Name of catchment (authority)	MoW, DWE	No problems	
FUNDER	Name of WP funder	VWC or printed on Water Point	Usually not legible anymore from WP. Many inconsistencies in spelling of names. Creates many more classes than actual exist (see 4.2.1)	Material (1a) Discursive (4a)
INSTALLER	Name of WP installer	VWC or printed on Water Point	Usually not legible anymore from WP. Many inconsistencies in spelling of names. Creates many more classes than actual exist (see 4.2.1)	Material (1a) Discursive (4a)
YEAR_OF_CO	Year of WP construction	VWC or printed on Water Point	Usually not legible anymore from WP. Many inconsistencies in spelling of names. Creates many more classes than actual exist (see 4.2.1)	Material (1a) Discursive (4a)
SOURCETYPE	Description of water source	DWE or WPC	No problems	
EXTRACTION	Type of extraction method	DWE or WPC	No problems	
WATERPOINT	Type of WP technology	DWE or WPC	No problems	
STATUS	Functionality of WP in 7 classes	WPC	Ambiguity in two of the 7 classes: Non-functional >3 m; Non-functional <6 m (see 4.2.6)	Material (1a) Observational (2b) Conceptual (3c) Discursive (4b) Discursive (4c)
STATUS2	Aggregation of STATUS in 2 classes	Geodata	Functional also includes WP that need repair. (see 4.2.6)	Material (1a) Material (1d) Conceptual (3c) Discursive (4d) (There is no evidence of fraud in the translation of STATUS into STATUS 2. It can however be seen as manipulation of data as STATUS 2 is an arbitrary rule based aggregation of STATUS)
BREAKDOWN_	Year of breakdown	VWC, Water User Group (WUG)	Inconsistencies exist where a breakdown year is recorded that precedes the construction year (see 4.2.3)	Material (1a) Discursive (4a) Discursive (4c)
HARDWARE_P	Hardware problem	VWC, DWE	No problems but the digital recording form contains several options more than the paper form (see 4.2.1)	Material (1a) Conceptual (3b)

Table A1. Cont.

REASON_WPT	Reason for hardware problem	WPC, VWC, DWE	Inconsistency in the interpretation of this attribute. Difficult to compare values (see 4.2.4)	Material (1a) Observational (2b) Conceptual (3a) Conceptual (3c) Discursive (4a) Discursive (4c)
WATER_QUAN	Water quantity of WP	VWC	No problems, though very subjective parameter (related to POPSERVED) (see 4.2.4)	Material (1a) Observational (2b) Conceptual (3a)
WATER_QUAL	Quality of WP water	WPC	Mix of objective testing & subjective judgement. (see 4.2.4)	Material (1c) Observational (2b) Discursive (4c)
SCHEME_MAN	Who is responsible for Water Scheme	VWC	No problems	
WP_MANAGEM	Who is responsible for WP	VWC	No problems	
WATER_PAYM	Whether and when payment is received for water	VWC	No problems	
AMOUNT_TSH	Amount of payment received for water use	VWC	This parameter can only be used in connection with WATER_PAYM as it has no unit of water use. Difficult to check consistency as value of 0 means either NO DATA AVAILABLE or AMOUNT_TSH is recorded with other WP in village to prevent double counting (see 4.2.3)	Material (1a) Conceptual (3a) Conceptual (3c) Discursive (4b) Discursive (4c)
PUBLIC_MEE	Whether a public meeting was held in the village about WP management	VWC	Ambiguity about the purpose of this meeting (see 4.2.3)	Discursive (4a) Discursive (4b) Discursive (4c)
GENERAL_CO	Comments made by the WPC regarding the WP in general	WPC	Very subjective parameter, difficult to classify. (see 4.2.4)	Material (1a) Observational (2b) Conceptual (3c) Discursive (4c)
GPS DATA	21 columns of automatic generated location parameters	Global Positioning System	Quality of data sensitive to human error	Material (1b) Observational (2a) Observational (2b)

References

1. Jiménez, A.; Pérez-Foguet, A. Challenges for Water Governance in Rural Water Supply: Lessons Learned from Tanzania. *Int. J. Water Resour. Dev.* **2010**, *26*, 235–248, doi:10.1080/07900621003775763.
2. Christopher Brown, J.; Purcell, M. There's nothing inherent about scale: political ecology, the local trap, and the politics of development in the Brazilian Amazon. *Geoforum* **2005**, *36*, 607–624, doi:10.1016/j.geoforum.2004.09.001.
3. Smoke, P.; Lewis, B. Fiscal decentralization in Indonesia: A new approach to an old idea. *World Dev.* **1996**, *24*, 1281–1299.
4. Smith, N. Homeless/global: scaling places. In *Mapping the futures: Local cultures, global change*; Bird, J., Ed.; Routledge: New York, NY, USA, 1993.
5. Ministry of Water and Irrigation. *The Water Supply and Sanitation Act, No. 12-2009*; Ministry of Water and Irrigation: Dar es Salaam, Tanzania, 2009.
6. Giné, R.; Pérez Foguet, A. Sustainability assessment of national rural water supply program in Tanzania. *Nat. Resour. Forum* **2008**, *32*, 327–342.
7. Welle, K. *Water Point Mapping in East Africa*; WaterAid: London, UK, 2010.
8. Ministry of Water and Livestock Development (MoWLD). *Water and Sanitation in Tanzania - Poverty Monitoring for the Sector Using National Surveys*; Government of Tanzania: Dodoma, Tanzania, 2002.
9. Welle, K. *Strategic review of WaterAid's water point mapping in East Africa Based on a review of Ethiopia, Tanzania, Kenya and Uganda*; WaterAid: London, UK, 2010.
10. Welle, K. *Learning for Advocacy and Good Practice—WaterAid Water Point Mapping, Report of Findings Based on Country Visits to Malawi and Tanzania*; WaterAid: London, UK, 2005.
11. United Republic of Tanzania (URT). *Water Sector Status Report 2009*; URT: Dar es Salaam, Tanzania, 2009.
12. Mwamwaja, S.A. *Mapping of Public Rural Water Service in Tanzania: A Case of Data Updating*; University of Twente: Enschede, The Netherlands, 2014.
13. Taylor, B. *Waterpoint Mapping, Planning and Obstacles to Equity in Rural Water Supply: A Review in Mpwapwa, Kongwa, Iramba and Nzega*; WaterAid: Dar es Salaam, Tanzania, 2009.
14. SNV Netherlands Development Organization. *Water Point Mapping: The Experience of SNV Tanzania*; SNV: The Hague, The Netherlands, 2010.
15. United Republic of Tanzania (URT). *Water Sector Status Report 2010*; URT: Dodoma, Tanzania, 2010.
16. United Republic of Tanzania (URT). *National Strategy for Growth and Reduction of Poverty II*; Ministry of Finance and Economic Affairs: Dar es Salaam, Tanzania, 2010.
17. United Republic of Tanzania (URT). *Water Sector Status Report 2013*; URT: Dar es Salaam, Tanzania, 2013.
18. Hutchings, M.T.; Dev, A.; Palaniappan, M.; Srinivasan, V.; Ramanathan, N.; Taylor, J.; Ross, N.; Luu, P. *mWASH: Mobile Phone Applications for the Water, Sanitation, and Hygiene Sector*; Pacific Institute: Oakland, CA, USA, 2012.
19. United Republic of Tanzania (URT). *Water Sector Status Report 2012*; URT: Dar es Salaam, Tanzania, 2012.
20. Kunz, W.; Rittel, H. W. J. Information science: On the structure of its problems. *Inf. Storage Retr.* **1972**, *8*, 95–98.
21. Rittel, H. W. J.; Webber, M. M. Dilemmas in a general theory of planning. *Policy Sci.* **1973**, *4*, 155–169.
22. Graham, J.P.; Hirai, M.; Kim, S. An Analysis of Water Collection Labor among Women and Children in 24 Sub-Saharan African Countries. *PLoS ONE* **2016**, *11*, 1–14, doi:10.1371/journal.pone.0155981.
23. Hemson, D. The Toughest of Chores: policy and practice in children collecting water in South Africa. *Policy Futures Educ.* **2007**, *5*, 315–326, doi:10.2304/pfie.2007.5.3.315.
24. Nganyanyuka, K. *Seeing Like a Citizen: Access to Water in Urban and Rural Tanzania*; University of Twente: Enschede, The Netherlands, 2017.
25. Rottenburg, R. *Far-Fetched Facts: A Parable of Development Aid*; MIT Press: Cambridge, MA, USA, 2009.
26. Pritchett, L.; Woolcock, M. Solutions when the solution is the problem: Arraying the disarray in development. *World Dev.* **2004**, *32*, 191–212, doi:10.1016/j.worlddev.2003.08.009.
27. World Bank. *World Development Report 2015: Mind and Society*; World Bank: Washington, DC, USA, 2014.
28. Taylor, J.R. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*; University Science Books: Herndon, VA, USA, 1997.
29. Narasimhan, S.; Jordache, C. *Data Reconciliation and Gross Error Detection*; Elsevier: Amsterdam, The Netherlands, 1999.
30. Florinsky, I.V. *Digital Terrain Analysis in Soil Science and Geology*; Elsevier: Amsterdam, The Netherlands, 2012.

31. Latifovic, R.; Pouliot, D.; Dillabaugh, C. Identification and correction of systematic error in NOAA AVHRR long-term satellite data record. *Remote Sens. Environ.* **2012**, *127*, 84–97, doi:10.1016/j.rse.2012.08.032.
32. Lintz, H. E.; Gray, A. N.; McCune, B. Effect of inventory method on niche models: Random versus systematic error. *Ecol. Inform.* **2013**, *18*, 20–34, doi:10.1016/j.ecoinf.2013.05.001.
33. Allchin, D. Error types. *Perspect. Sci.* **2001**, *9*, 1–16.
34. National Bureau of Statistics (NBS). *2012 Population and Housing Census*; NBS: Dar es Salaam, Tanzania, 2013.
35. United Republic of Tanzania (URT). *Big Results Now Brings Clean Water Supply to 752,000 Villagers in Three Months*; URT: Dar es Salaam, Tanzania, 2013.
36. Giné-Garriga, R.; Jiménez-Fernández de Palencia, A.; Pérez-Foguet, A. Water-sanitation-hygiene mapping: an improved approach for data collection at local level. *Sci. Total Environ.* **2013**, *463–464*, 700–711, doi:10.1016/j.scitotenv.2013.06.005.
37. Horn, R. E.; Weber, R. P. *New Tools For Resolving Wicked Problems: Mess Mapping and Resolution Mapping Processes*; Available online: http://www.strategykinetics.com/New_Tools_For_Resolving_Wicked_Problems.pdf (accessed on 3 August 2017).
38. Global Water Challenge. Available online: www.globalwaterchallenge.org (accessed on 4 August 2017).



© 2017 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).