

# The Wananchi Survey in Tanzania

## Concept Note, January 2012

### 1. Background

Information about the activities, preferences and aspirations, opinions, experiences and welfare of citizens in Tanzania is collected mostly through large scale household surveys. Because these surveys are costly, they are implemented infrequently. This limits the ability to monitor change and reduces citizens' ability to hold authorities to account. It also limits the incentives for authorities to adjust their actions in light of survey findings.

For example, the 2007 Household Budget Survey (HBS) found that since 2001, poverty had declined very little, suggesting that previous poverty reducing policies were unsuccessful. If in response to this, the Government of Tanzania implements a different set of policies and programs and the frequency of household surveys is not increased, it could take another 6 years until the effects of these interventions are known. In such an environment, myopic decision makers, or decision makers who are uncertain about the impact of their interventions, may decide not to change anything at all.

This concept note outlines an approach to data gathering that combines the strength of household surveys (representativeness) with possibilities offered by mobile phones (low cost, high frequency feedback). This approach, called the *Wananchi Survey*<sup>1</sup>, aims to collect data at a fraction of the cost of ordinary household surveys, in a way that is more frequent and more responsive to changing data needs.

### 2. Wananchi Survey, the missing link

In Tanzania timely information on the perspectives and experiences of citizens is unavailable. What is available are summary reports from a number of household surveys which collect information on a range of topics: poverty (HBS), labor (LFS), health (DHS), agriculture (Agriculture Census), accountability (Afrobarometer), people's perspectives (VoP), welfare (NPS) and politics (REDET). Each of these surveys is implemented after an extensive planning cycle. A typical survey takes a year in preparation, at least a few months in data collection and then another 6-12 months until analyses are done. As a consequence the data provide excellent descriptive statistics, but they are never very well-timed for evaluation of policies.

These aspects of the existing data landscape leave two gaps, as there is a desire to:

- (i) ensure that more data is in the public domain; and

---

<sup>1</sup> "*Wananchi*" means citizens in Kiswahili.

- (ii) Regularly have information on time-sensitive issues, such as drought and food availability, people’s opinions about governance, the quality of public service delivery, or citizens’ ability to exercise agency.

Such information could be useful to a diverse audience including policy makers and implementers, members of parliament, newspapers, and donors, as well as *Twaweza*<sup>2</sup>. To fill the existing gap, *Uwazi* intends to create a citizen-based data collection system, the Wananchi Survey. Reasons for *Uwazi* to take on the Wananchi Survey include:

- *Uwazi* is independent and autonomous with respect to what data to collect, how to analyze it and where to publicize it;
- *Uwazi* is committed to quality data and analysis;
- *Uwazi* is committed to making all of its (anonymized) data publicly available free of charge;
- *Uwazi* has in-house analysts who can produce policy briefs in a short period of time;
- *Uwazi* has the capability to make analyses and data available in a variety of ways to journalists, parliamentarians, researchers, decisions makers and other interested parties including through its website and policy briefs and intends to produce short, accessible, reports of 2-4 pages for each topic on which data is collected.

### 3. Approach

Ownership and use of mobile phones is growing rapidly in Tanzania. Evidence from the HBS 2007 survey (Table 1) demonstrates that about 25 percent of households own at least one phone, with ownership being highest in Dar es Salaam (almost 66 percent) and lowest in rural areas (14 percent). A 2010 Gallup survey of mobile phone ownership across 17 African countries found that phone ownership had increased since 2007: 35% of *adults* in Tanzania own mobile phones. This implied a household access rate higher than 35%, since presumably some of the 65% of adults who do not own phones themselves live with spouses or other relatives who do own phones.<sup>3</sup> This impression was confirmed and updated by the 2010 DHS, which found that 46% of households owned phones, and the 2010 Uwezo survey, which found a similar percentage (48%) of phone ownership. Similarly, a pilot survey by *Uwazi* in Dar es Salaam during 2010 suggested that household phone ownership (in Dar es Salaam) had increased since the 2007 HBS, and was between 84 percent–90 percent in 2010.

---

<sup>2</sup> Twaweza identifies information as key to making change happen: “*When exposed to ferment of information and ideas, and practical tools or pathways in which to turn these ideas into actions, citizens can become drivers of their own development and act as co-creators of democracy*” (Twaweza 2008: p.18) –italics in the original.

<sup>3</sup> See report at <http://www.gallup.com/poll/149519/mobile-phone-access-varies-widely-sub-saharan-africa.aspx>

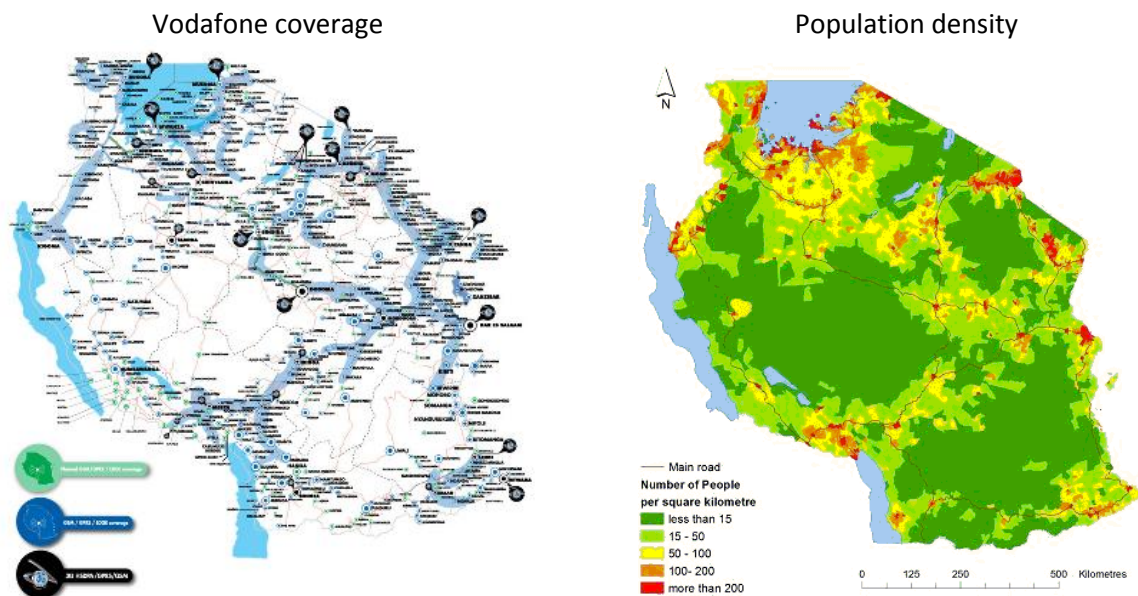
**Table 1: Percent of households owning at least one mobile phone, by area (2007 HBS)**

Domain	Percent of households owning at least one mobile phone
Dar es Salaam	65.8
Other urban areas	42.5
Rural areas	13.9
Tanzania	24.9

Source: *Uwazi*, based on HBS, 2007

Not only is the ownership of mobile phones widespread (and expanding rapidly), phones are available at relatively low cost (Tshs 40,000 or less) while networks cover most of the population (the areas not covered are of low population density—Figure 1).

**Figure 1: Mobile phone coverage and population density**



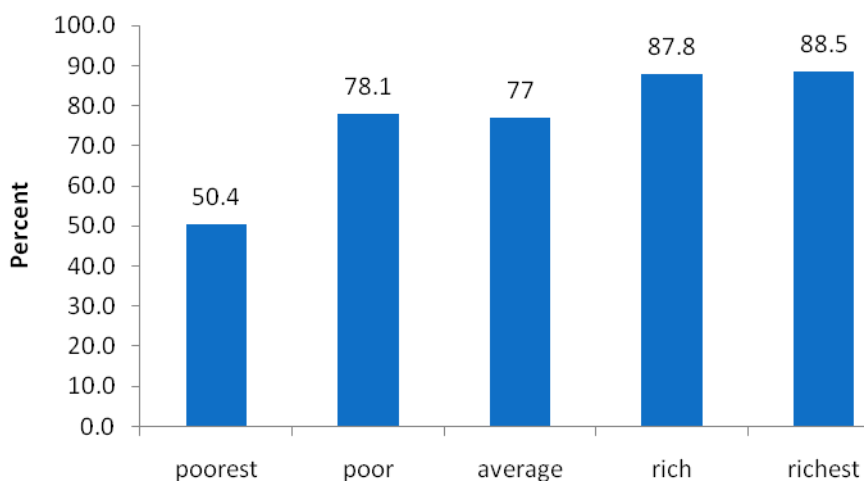
Source: Vodacom & PHDR 2007.

The expansion of access to mobile phones and potential to use them to reach the majority of population creates new opportunities for data collection. Through the use of the mobile phones, it is now imaginable to collect representative data at the cost of a few SMSes or short phone calls to a reasonably sized household panel. Twaweza/Uwazi has piloted this approach in Dar es Salaam, and has shown that high frequency data collection is possible by way of a short phone call. Wananchi survey will

implement this approach on a larger scale in Tanzania starting in late 2011. The survey shall be conducted once every month, and if successful shall be replicated it in all EAC countries of Twaweza focus in the coming years.

One challenge with this approach, however, is that households owning a mobile phone may not be representative of the entire population (see Table 2). Analysis of the HBS shows that members of households that are wealthier, better educated and live in better circumstances are more likely to own phones. This holds in rural areas, and equally in Dar es Salaam. Similar patterns were identified in the Uwazi Dar es Salaam pilot survey. For example, as Figure 2 shows, only 50% of households in the lowest income group in Dar es Salaam own a phone, while 89% of those in the highest group own a phone.

***Figure 2: phone ownership by wealth quintile in Dar es Salaam***



Source: Uwazi, based on Wananchi Survey pilot in Dar es Salaam (2010)

As a consequence the Wananchi Survey shall ensure foremost that all respondents recruited for the survey have access to a mobile phone, either a phone provided for free by the recruiting team or one that they already own. This will eliminate bias in the sample<sup>4</sup> caused by the fact that poorer households do not own phones but also may be practical since

---

<sup>4</sup> Attrition is the problem of cohort members preferring no longer to be interviewed, and dropping out, or moving and being unable to be traced. It can lead to biased results if those who drop out share common characteristics (e.g. low income, low basic skills).

<http://www.cls.ioe.ac.uk/text.asp?section=00010001000500150009>

**Table 2: Social economic characteristics of households by mobile phone ownership**

	Dar es Salaam		Rural Areas	
	With Phone	Without Phones	With Phone	Without Phones
Age of the head of household (Years)	39.9 (0.24)	40.4 (0.43)	43.1 (0.67)	46.5 (0.30)
Education by head (Years)	9.8 (0.10)	6.3 (0.12)	7.4 (0.20)	4.5 (0.08)
Family size (family members)	3.9 (0.05)	3.3 (0.07)	5.6 (0.17)	4.9 (0.05)
Access to clean water* (Percent)	64 (0.99)	58 (1.48)	30 (2.16)	22 (0.77)
Owning a television (Percent)	55 (1.03)	11 (0.93)	12.5 (1.56)	0.2 (0.07)
Owning a radio (Percent)	84 (0.75)	65 (1.43)	92 (1.29)	57 (0.93)
Owning a bicycle (Percent)	14 (0.72)	9 (0.87)	65 (2.25)	42 (0.92)
House has strong walls** (Percent)	95 (0.47)	82 (1.16)	41 (2.33)	18 (0.71)
House has strong roof*** (Percent)	98 (0.25)	95 (0.68)	72 (2.12)	33 (0.88)
Poorest Quintile (Percent)	15 (0.74)	29 (1.36)	6 (1.11)	22 (0.78)
2 <sup>nd</sup> Quintile (Percent)	18 (0.80)	23 (1.26)	11 (1.45)	22 (0.77)
3 <sup>rd</sup> Quintile (Percent)	20 (0.83)	20 (1.19)	19 (1.84)	20 (0.75)
4 <sup>th</sup> Quintile (Percent)	22 (0.85)	16 (1.11)	28 (2.11)	19 (0.73)
Wealthiest Quintile (Percent)	25 (0.89)	12 (0.97)	38 (2.29)	17 (0.71)

(Standard errors in parentheses)

\* Clean water includes: Private piped (tap) water in housing unit, private piped (tap) water outside housing unit, piped water on neighbor's housing unit and Piped water on community supply

\*\* Strong walls include those made from baked / burnt bricks, concrete, cement and stones

\*\*\* Strong roofs include those made of concrete, cement, metal sheets, asbestos and tiles

Source: *Uwazi-Twaweza*, based on HBS, 2007

#### 4. Sample Design

The survey relies on a standard sampling procedure (EA/village/city block sample selection; listing; household selection) to establish a baseline sample that shall be followed up using mobile phones. The aim is to create a sample representative of two domains: rural and urban. The proposed design of the survey is therefore:

- (i) A baseline survey
- (ii) All respondents recruited for the survey are provided a mobile phone
- (iii) Additional data is collected using the mobile phone provided.

To determine the sample size, we calculate the number of clusters (EAs) necessary to arrive at a predetermined *confidence interval*. For our purposes, a confidence interval represents an interval around the survey finding for which we can be 95% certain that it contains the true (or population) value. For example, suppose the Wananchi Survey finds that 50 percent of households favor Kikwete as next president. If the confidence interval is +/- 5 percentage points we are confident (95% certain) that population estimate – i.e. the number we would find if we surveyed the full population – is between 45 and 55 percent. A confidence interval of 5 percent or less is what we aim for here.

Using this precision target one can calculate the required number of EAs (clusters). In this calculation we take into account the clustered nature of the survey (see e.g. Bennett, 1991, for the formulae involved). We assume that 10 households are sampled per EA. An important parameter is how “alike” EAs are: the more homogeneous the EAs, the more of them you need to say something about the population mean.

For aggregate proportions, say the nationwide percentage of the population intending to vote for candidate X or having access to an improved water source, we find that with 10 hh per EA and high (low) EA homogeneity the required number of EAs is 148 (67). Since we do not know all the questions beforehand so cannot judge the level of homogeneity we choose for a sample of 150 EAs.

There are two further reasons for not choosing the low number (of 67 clusters). The precision of the estimate will decrease when households drop out of the survey (attrition); for example, when the number of respondents remaining in each EA is lowered to 6 the confidence interval widens to 10 and 7 percent for nationwide estimates. A second argument is that it should be possible to present estimates for certain subgroups, e.g. the bottom or top asset quintile. In this case the confidence interval widens further.

While a large sample size offers many possibilities, there are other aspects to consider as well. The survey will be the first largest mobile phone based household level survey in Tanzania, and managing it with a very large sample may prove challenging. For this reason, a sample size of 150 EAs with 10 respondents each is proposed.

## 5. Sample Selection

The Wananchi Survey (WS) will use as a sampling frame the sample of EAs and households from the Uwezo Learning Assessment Survey. In Tanzania, Kenya and Uganda, Uwezo follows a common sampling procedure. The numbers of households in these surveys are large: in each district in these three countries, 30 EAs are sampled and in each EA 20 households. The household numbers sampled by Uwezo are thus about 80,000 (TZ, 2011), 73,000 (KE, 2010) and 50,000 (UG, 2011). The 1500 households sampled for WS will thus constitute only a small fraction (2-3 percent) of the Uwezo sample.

The Uwezo surveys follow a standard 2-stage sampling procedure. The frame for the first stage of selection is a list of Enumeration Areas based on Census data; EAs are the primary sampling unit for Uwezo. Uwezo selects (randomly) 30 EAs with PPS in each district. The frame for the second stage of (household) selection is an updated list of households obtained from a listing exercise in each selected EA.

The use of the Uwezo sample as a sampling frame is attractive for a number of reasons. First, it presents a recent list of households for a large set of EAs. Since EAs have been sampled randomly (by statisticians from the respective National Bureaus of Statistics) with PPS, full listing has been carried out and random household sampling was carried out, the Uwezo sample can be used as a sampling frame to select a nationally representative sample for Wananchi. The WS sampling procedure should select 150 EAs with PPS from the Uwezo sample of EAs first; then select 10 households out of the 20 in each EA. The recommendation is to create EA wealth deciles first and select one household from each decile, while keeping the second as a possible replacement.

A second advantage of using the Uwezo sample is that it allows for a replication of the sampling approach in all three countries. Since the share of households selected for WS is so small, the interference with Uwezo operations should be negligible. In any case, logistics will be organized such that optimal synergy is realized, e.g. WS follows after the Uwezo assessment has been completed and solicits cooperation from Uwezo district and EA coordinators.

In concrete terms the procedure will work as follows: Uwezo EA sampling will be completed in December 2012. Uwezo household listing takes place in January 2012 and household sampling follows early February. Once the Uwezo data have been collected (February – April 2012) and entered (data entry completion expected ...?) the data are shared with Uwazi. Uwazi then proceeds with EA / household sampling. EA sampling by WS could already take place early January if this would facilitate logistics later in the year.

## 6. Non response, attrition and cohort effects

Non response can occur in three instances:

1. Refusal to participate in the interview during (*re*)visit;
2. At the end of the interview household indicates not to be interested in follow up;

### 3. During follow up.

The first two instances can be treated as typical household non-response and as it occurs during the first (re)visit a new household can be drawn from the listing. Information collected during the listing and during the interview can be used to correct for this type of non-response. The third type of non-response occurs during the follow up interviews and needs sharper strategies to address as its effect on the desired overall sample of the wananchi survey can be more detrimental.

As the Dar es Salaam pilot has shown attrition during follow up can be very high. In the Dar es Salaam pilot attrition rate during follow up was estimated at 35%<sup>5</sup>. This happened due to several reasons: wrongly recorded phone numbers during the baseline interview; people changing phone numbers due to loss or other reasons, migrations and loss of interest in the survey.

This suggests that a target sample of 1500 (10 people per enumeration area) may turn out in actual fact to be an effective sample of 975 respondents. In this case the error margins will go up and may limit possibility of presenting results with detailed breakdown by rural urban sub-groups and wealth quintiles within the sub-groups.

Care will be taken during (re)visits to ensure that numbers are correctly recorded; and that extensive training and discussion with potential respondents takes place to clear doubts and address concerns ahead of the survey. On the other hand occurrence of the third type of attrition may also depend on the incentives offered; and rapport that is established between respondents and call centre operators. In the Dar es Salaam pilot for example, after the initial attrition in the first weeks of follow up, the response rate stabilized and it could be partly linked to the fact that the same person(s) called the interviewees each week. Experience from the Dar es Salaam survey suggests also that response rates to a single call with rewards are approximately 65%. That the survey was implemented over a short period (20 weeks) and may not provide a realistic estimate of attrition over the long run needs to be borne in mind, however.

It can be anticipated that with well-designed incentives (assistance for phone charging—solar chargers or vouchers to use for charging phones in local kiosks; follow up by SMS or voice calls; phone credit; free SMSes; flexibility in the times at which the household is called; lotteries; notifications on what has been done with the data etc.), and proper explanations and training of respondents prior to the start of the survey the rate of non-response can be reduced. Also, since background data on wealth quintile from the baseline survey exists, a further option could be to offer poor households higher rewards for participation. Another strategy to reduce attrition is to collect mobile phone number of neighbors, relatives or friends, to enable tracking of respondents that drop out of the survey. This will need to be tested. If non-response is high, it will have an impact on the size of the sample.

---

<sup>5</sup> The sample was 550 respondents. Of these 90 did not have a phone and therefore could not have been expected to participate. Of the remaining 460 155 dropped out of the survey. The respondents stabilised at 305 in the last 10 weeks of the survey.

It is important to note also that even with properly designed incentives and training, some respondents may still drop out. A similar survey—the CORDAID survey conducted in Southern Sudan—and which incorporated incentives that are similar to the ones being considered for the Wananchi survey had a very high attrition rate half a year into the survey<sup>6</sup>.

This suggests that adjusting for non-participation ahead of the survey is essential to avoid unpleasant surprises. We propose a protocol for household replacement based on a pre-defined threshold which can be based on confidence interval calculations; e.g. sample replacements in each village when the number of respondents falls below 6. Baseline data could be used to select replacements on observable characteristics.

And finally in the worst case scenario, i.e. if all strategies fail and the sample is reduced and or biased towards certain type of households, prior to analysis survey data can be re-weighted to account for differential rates of attrition. This increases standard errors but is a reasonable response to the challenge of differential attrition. <sup>7</sup>

Cohort effects occur because the respondent households age over time. This is an issue to be considered, but not till well into the first (and possibly second) year of data collection. One possibility is to annually rotate a share of the sample, though there would equally be advantages to keeping part of the sample fixed over time.

## 7. What data to collect

The baseline survey shall normally comprise a household questionnaire, a community questionnaire, a school questionnaire, and a health facility questionnaire. Their content is described in Table 2.

*Table 2: Outline of baseline household questionnaire*

Category	Questions envisaged
Basic information	Identifiers, replacement household, enumerator details
Household roster	Name, age, gender, years of education completed, literacy
Education	School currently attended, distance to school, type of school
Employment	Place of work, main and secondary occupation
Health	Health facility frequented, distance to facility
Water	Water source used
Assets	Radio, television, bicycle, watch, mobile phone
Housing	Housing conditions
Media-voice-agency	Sources of information, participation in meetings, where do you go to solve problems, who listens to you
Mobile phone	Has phone; willingness to participate in survey
Contact details	Mobile phone contacts of relatives, neighbors, friends

<sup>6</sup> The CORDAID mobile phone survey panel started in December 2010 with a response rate of 70%. Six months later, in June 2011, the response rate had dropped to 48%. (*Unofficial lessons learnt note, shared by Johannes Von Engelhardt in October 2011*)

<sup>7</sup> Attrition may also occur if sim cards are not registered. Registration of sim cards is obligatory and *Uwazi* will have to register phones that are provided, or enable respondents to do so.

---

Source: *Uwazi*

To the extent information is available from the Uwezo data the amount of information to be collected will be reduced and the length of the questionnaire shortened. For example, community, school and health facility questionnaires may not be necessary so only a household questionnaire will be administered. Following the (re)visit, phone calls from a call center will be used to collect additional data.

It is envisaged that such data is collected on a monthly basis. The data collected will typically be specific, often only requiring a yes/no response or a response from a list provided. For instance, prior to elections the call center enumerator could ask (i) whether respondent registered to vote or (ii) which presidential candidate plans to support. Prior to a Parliament meeting enumerators may ask (i) Whether the respondent is aware of the upcoming meeting and (ii) What issues they expect the Legislators to address in the session. More complicated questions could also be asked, for instance, about whether yesterday the oldest child attended school. The latter question could then be followed up with other questions including why the child did not go and/or whether the teacher was absent.

In some instances, the survey will be targeted to a subset of the sample, for example a question like “Do you have water at your home today?” may have to be adapted to different sub-samples. Households with access to tap water would receive a question whether they have access to tap water; household with access to well water would receive questions relevant about the well (waiting time, for instance).

## 8. Reporting

Once collected *Uwazi* will use the information to prepare brief reports (2-4 pages), which it will put out for use by journalists, parliamentarians, decisions makers, researchers, and other interested parties. Hence, if through mobile surveys information was collected on (i) which water source people usually use, (ii) whether this source was functional last week and (iii) what the combined waiting and travel time was to collect water, then a report will be produced presenting:

- Information (%) on type of water sources, whether sources are functional and average travel and waiting time for rural and urban areas;
- Percent changes in these variables since the *baseline (or revisit) survey*
- Information on these variables broken down by levels of education of the head of household, household wealth or distance of the village to e.g. the district centre.

The information would be presented in an accessible format (graphics; maps) and may be complemented with information from secondary sources on e.g. public spending on water; relation between water-borne diseases and access to clean water or changes in water tariffs. Case studies presenting information collected by having a longer phone conversation with selected respondents could also be reported.

## 9. Quality control

Data quality is an extremely important consideration. If data collected from the Wananchi survey is meant to inform policy, then accuracy is a very high priority.

It is understood that potential data quality issues have been identified in relation to the baseline survey. If these data quality issues are limited to data entry problems, two potential solutions suggest themselves. The first would be to introduce greater Twaweza/Uwazi supervision into the data entry process. Uwazi expects to recruit a Wananchi survey manager with household survey knowledge and experience. Informed by close consultation with the external evaluation teams, the survey manager could supervise data entry by the survey firm more closely than was the case during the baseline survey. Second, Twaweza/Uwazi can write a contract with the survey firm which conditions payment on achievement of accuracy targets with respect to data entry.

Another important issue (observed during the baseline) was that enumerator morale suffered due to late payment by the survey firm. Enumerators that are not promptly paid are much less likely to produce quality data. Twaweza could insist contractually on being provided evidence that enumerators have been paid promptly (for example evidence that they receive their DSA money in advance, and half of their fee at a mid-point during fieldwork.)

A third issue relates to questionnaire length. The baseline questionnaire was extremely long, which may have affected data quality. This will almost certainly not be an issue in the mobile phone survey.

A fourth issue relates to training. For certain aspects of the questionnaire, it was not possible to give enumerators the kind of specialized training that would have been ideal. For example, many of the questions on the health facility questionnaire related to somewhat technical issues (requiring familiarity of drug names, or types of medical equipment), and specialized training for team leaders would have been helpful. This again should not be an issue with the wananchi survey, which will most likely not contain facility-related questions.

A final issue relates to back checks by supervisors. In the original baseline survey plan, provision was made for extensive back checks. However, these are costly for the survey team to implement, since they require extensive travel and time in the field by supervisors, and it is possible that they did not implement them. More generally, writing a contract where payment is contingent on 3<sup>rd</sup> party evaluation of data quality or achievement of certain data quality targets as specified and determined by Twaweza, might be a reasonable step.

Several other broader quality control measures could be adopted during the follow up mobile phone interviews as follows. The ability to speak to respondents directly by calling them creates possibilities for quality control not available to traditional household surveys. Call centre operators could be asked to call a selection of respondents (say 50) to check whether questions have been properly understood and to verify whether errors have been made. Some of these respondents will be selected randomly; others could be called because their responses are flagged by the consistency checks which will have to be built

into the data collection system. This could easily be something that Twaweza implements internally, in order to provide some quality control over the mobile survey.

## **10. Data access**

All information from the Wananchi Survey will be anonymized and made available to the public through the *Uwazi* website where the data can be downloaded in raw form for further analysis. It is envisaged that immediately after *Uwazi* has analyzed and published, the data is put in the public domain and made accessible free of charge. *Uwazi* is committed to putting its data in the public domain within several weeks of collection.

## **11. Other considerations**

- Registration of sim cards is obligatory and *Uwazi-Twaweza* will have to register phones that are provided, or enable respondents to do so.
- Providing respondents with mobile phones will affect the representativeness of the sample in dimensions such as ownership of mobile phones, access to information or money transfers.