

Sauti za Wananchi data release notes

October 2013

General information

In accordance with Twaweza's commitment to open data, the *Sauti za Wananchi* mobile phone data are made public through the website. This note outlines the nature of the data release and details which files and documents are provided.

The data and files made available here remain the property of Twaweza and are provided for non-profit, research purposes only. We advise all users of the data to read the data use agreement found below; downloading of the data implies agreeing to the requirements. Please note that all data is provided in raw form and on an "as is" basis. This means Twaweza provides absolutely no warranty as to the underlying data quality or its fitness for purpose. We also cannot provide any form of follow-up service to this data release, and therefore will generally not reply to specific questions from individual researchers or organizations as regards the origin or uses of the data.

To facilitate public use of the data, a minimal verification process has been applied. Design weights are provided with the household data and the calculation of weights is described in a separate file named "SzW baseline weights note". Again, Twaweza does not provide any warranty as to the quality of the weights or their fitness for purpose.

For further information regarding the *Sauti za Wananchi* data collection process, and for reports that employ these data, please refer to the list of publications found at: <http://www.twaweza.org/go/sauti-za-wananchi-english>.

Data file information

Two types of SzW data are released through the Twaweza website <http://www.twaweza.org/go/sauti-za-wananchi-english> : baseline data from the 2012 Pen and Paper Interview (PAPI) phase and data from the subsequent Computer Aided Telephone Interviews (CATI) phase. CATI data sets are added monthly.

Sauti za Wananchi Baseline data

Household data		
<i>Filename</i>	<i>Description / Notes</i>	<i>Format</i>
SzW.data.hh.baseline.2012.dta	Individual- and household-level data file -Includes 400 reserve households -Includes design weights	Stata v12.1
SzW.datamap.hh.baseline.2012.xlsx	Data map with variable / value labels	Excel
SzW.qnr.hh.baseline.2012.pdf	Questionnaire	Pdf

Community data		
<i>Filename</i>	<i>Description / Notes</i>	<i>Format</i>
SzW.data.comm.baseline.2012.dta	Community (EA)-level data file	Stata v12.1
SzW.datamap.comm.baseline.2012.xlsx	Data map with variable / value labels	Excel
SzW.qnr.comm.baseline.2012.pdf	Questionnaire	Pdf

School data		
<i>Filename</i>	<i>Description / Notes</i>	<i>Format</i>
SzW.data.sch.baseline.2012.dta	School-level data file	Stata v12.1
SzW.datamap.sch.baseline.2012.xlsx	Data map with variable / value labels	Excel
SzW.qnr.sch.baseline.2012.pdf	Questionnaire	Pdf

Health facility data		
<i>Filename</i>	<i>Description / Notes</i>	<i>Format</i>
SzW.data.hf.baseline.2012.dta	Health facility-level data file	Stata v12.1
SzW.datamap.hf.baseline.2012.xlsx	Data map with variable / value labels	Excel
SzW.qnr.hf.baseline.2012.pdf	Questionnaire	Pdf

All datasets have been anonymized by removing low level geographic identifiers, names, phone numbers and so forth. For this reason, some questions/variables in the questionnaire are not found in the datasets.

Sauti za Wananchi CATI data

CATI data are added periodically to the Twaweza website. We therefore provide the general file naming conventions and document types provided. There are five different types of CATI data, each with their own respondent type. The corresponding labels in the file name are:

Data type	Label
Household data	hh
School data	sch
Health facility data	hf
Citizen monitoring data	cm
Water management system	wms

For each dataset typically three files are provided: (1) the dataset itself (Stata with .dta extension); (2) the datamap or codebook which contains all variable names and value labels used (Excel); and (3) the questionnaire corresponding to the dataset (pdf).

Preliminary CATI waves: a special feature of the Sauti CATI phase consists of the three so-called preliminary waves. These CATI waves were completed in the period between end of the baseline PAPI phase (when baseline data processing was ongoing) and the official launch of the call phase. These preliminary waves were considered crucial to keep contact with the respondents after the field visit and phone distribution. The data collected have the special label “BLPW” (Baseline Preliminary Wave).

Merging instruction

The SzW dataset contains 200 **enumeration areas** or EAs, with unique EA numbers 1 to 200. The variable containing the EA number is called “ea”. All EA level datasets (community, school, facility, citizen monitoring and water management) should contain and can be merged using the variable “ea”.

Each **household** in the SzW dataset is uniquely identified by the variable “uhn” or Unique Household Number. Each uhn consists of the unique ea code to which are added two digits to identify the household. In the baseline PAPI data the uhn runs from 101 (ea = 1, hh = 01) to 20012 (ea = 200, hh=12). Note that in each EA, the household numbers 11 and 12 were used for reserve households.

All household level datasets should contain and can be merged using the variable “uhn”. In case it is missing, the ea code can easily be derived from the uhn by dividing by 100 and rounding.

Data use agreement

By the act of downloading and employing this data, the user agrees to the following:

- *Sauti za Wananchi* data are freely provided on an 'as is' basis, for the purpose of legitimate non-profit research only.
- Twaweza / *Sauti za Wananchi* provides no warranty, expressed or implied, as to the accuracy, reliability or completeness of this data.
- Twaweza / *Sauti za Wananchi* bears no responsibility for uses of the data by third parties, or for interpretations or inferences based upon such uses. Any findings and conclusions derived from these data are solely attributable to the user and do not reflect those of Twaweza / *Sauti za Wananchi*.
- The user will not make available, transmit, copy or reproduce the data in any form to third parties without the prior express consent of Twaweza / *Sauti za Wananchi*.
- Twaweza / *Sauti za Wananchi* prohibits use of this data for the purpose of identifying and/or reporting results at the level of any specific individual, school, facility or local community (that is, the primary sampling units). If the identity of any individual, school, facility or community described in the data set is discovered inadvertently, the user will make no disclosure or other use of this information and kindly inform the survey management by sending an email to uwazi@twaweza.org.
- *Sauti za Wananchi* data are protected by copyright. Authors of any academic papers, reports or other published work based on this data are required to acknowledge *Sauti za Wananchi* and Twaweza in the text as follows:
 - For papers, reports, other output: "The findings in this brief (report /output) are based on data collected by *Sauti za Wananchi*, Africa's first nationally representative mobile phone survey fielded by Twaweza. The findings, interpretations, and conclusions expressed in this brief/report/output are entirely those of the authors and do not necessarily represent the views of Twaweza." This acknowledgement should be in a prominent place.
 - For stand-alone graphs, tables, stand-alone facts: "Twaweza, SzW survey (Round[n]), (Month), (Year)" below the graph, table, fact.
- Users are kindly requested to inform Twaweza and the Uwazi research unit of any publication based on *Sauti za Wananchi* data by sending an email to uwazi@twaweza.org. Please enclose a soft copy of the report or publication.