**Project Information**

We are researching off-grid sanitation in Kenya, Peru, Haiti, and South Africa, focusing on how different user groups access and utilize container-based sanitation. Participants, acting as citizen scientists, will collect confidential data through a dedicated app. The project aims to identify barriers and issues, sharing insights with other cities interested in implementing container-based sanitation for off-grid areas. The 12-month project offers participants small incentives for weekly engagement, requiring random selection, attendance at a training workshop, and access to a smartphone with the necessary tools. We conducted weekly smartphone surveys in over 300 households across informal settlements in Kenya, South Africa, and Peru over the 12-month period. These surveys aimed to understand human-environment interactions by capturing daily life, wellbeing, income, and socioeconomic variables at a weekly resolution, contributing to more informed analyses and decision-making.

This work took place under the ‘Scaling-up Off-grid Sanitation’ project (SOS; ES/T007877/1), funded with support from the United Kingdom’s Global Challenge Research Fund, via the Economic and Social Research Council. This research involves researchers in the UK, USA, Haiti, Kenya, Peru and South Africa. The project is led by Cranfield University (UK), but the survey itself is being led by Dr. Simon Wilcock at Bangor University (UK) and can be contacted at s.willcock@bangor.ac.uk.

In total, 310 households were included in our final data set and the survey was live from 13/12/21 to 15/07/23. The survey was live for 80 weeks due to the fact that there was a lag time both within and between countries recruiting participants. Participants were able to leave the project after 52 weeks of participation.

**This folder contains**

1. **Folder: SOS\_Consent and training materials**
2. SOS\_Permits

In addition to the ethical approval (below IV) additional permits were acquired for in country research. These are saved as pdf files.

1. Privacy Policy for Data exchange.docx

Privacy policy for using our Data Exchange app.

1. SOS\_Consent\_form.docx

Participant consent form in English, used across all countries.

1. SOS Ethics approval letter March 2021.docx

Ethical approval letter from Bangor University.

1. SOS\_Phone\_setup\_documentation.docx

Documentation describing the app installation and download of the odk surveys onto a phone.

1. SOS\_SamplingStrategy.pdf

Documentation describing the sampling strategy.

1. SOS\_ Survey\_Design\_References.docx

Table describing key indexes or standardised questionnaires within the survey with references.

1. SOS\_Training\_2021\_english.pptx

Training materials provided to the in-country teams to assist with training workshops.

1. **Folder: SOS\_Datasets**
	* + 1. SOS\_Consolidated Users.xlsx

Consolidated final users within the SOS project and finalised according to the SOS\_Cleaning\_Protocol in section 3.i below. This is the master data set allowing the link between demographic variables, as well as country and CBS user type to the ID\_Key. This ID is unique to an individual, and all data sets are linked to this ID.

All data sets below have been cleaned and anonymised according to the SOS\_Cleaning protocol.docx in section 3.i below. All data for all countries have been merged into a single file by task topic. All variable names relate to the ODK codes which can be found in Folder 4. This also describes the data types (e.g. single or multiple choice).

* + - 1. SOS\_Demographics Monthly.xlsx
			2. SOS\_Demographics Weekly .xlsx
			3. SOS\_Elections.xlsx
			4. SOS\_Electricity.xlsx
			5. SOS\_Poverty\_likelihoods\_Kenya.xlsx
			6. SOS\_Poverty\_likelihoods\_Peru.xlsx
			7. SOS\_Poverty\_likelihoods\_SA.xlsx
			8. SOS\_Sanitation Access.xlsx
			9. SOS\_Shocks.xlsx
			10. SOS\_SNA.xlsx
			11. SOS\_WASH Monthly.xlsx
			12. SOS\_WASH Once.xlsx
			13. SOS\_WASH Quarterly.xlsx
			14. SOS\_WASH Weekly.xlsx
			15. SOS\_Wellbeing.xlsx
1. **Folder: SOS\_Survey release dates and download code**
	* + 1. SOS\_Cleaning protocol.docx

Comprehensive document describing the cleaning protocol for all data sets in section 2.

* + - 1. SOS\_Compensation\_Calculator.xlsx

Calculator used to calculate scores per task that each respondent could receive if they were using their own phone or a project phone. This was localised by country and dependant on phone and data costs.

* + - 1. SOS\_Points\_assignment.R

R code Assigning points to individuals based on weekly scores.

* + - 1. SOS\_Server data download.R

R code accessing the central server (now deleted) and merging the points scored by each individual by each week.

* + - 1. SOS\_Survey\_release\_dates.xlsx

An excel file listing the names of each ODK form (see all forms in section 4) and their release dates and points value. Baseline surveys were open ended, weekly surveys open for 7 calendar days. Some tasks had bonus sections.

* + - 1. SOS\_Total\_Form.list.xlsx

List of all 3022 forms uploaded to the ODK central server.

* + - 1. SOS\_Python Code (Folder of 15 files).

This Folder contains 15 Python code files which clean each of the individual data sets. Each file is named according to the data set in question; for example Python\_code\_Wellbeing\_Dataset.ipynb is the code used to clean the wellbeing data sets. These were developed by Ana Casas, School of Water, Energy and Environment, Cranfield University.

1. **Folder: SOS\_Survey\_ODK\_code**

This folder contains all surveys uploaded to the ODK central server. Each ODK form is attached to a “prefill.xlsx” file. This contains information on the points and longevity of each form. Within the ODK form itself there are three tabs “Survey”, “choices” and “settings”. The survey tab is all possible questions that the respondent could answer. See “label::en” for English questions and “label::XX” for alternative languages. Each question is related to a single variable “name”, regardless of the language presented to the respondent. Column header “type” refers to the type of question, this may be free text, multiple choice or GPS position. See <https://getodk.org/> for all code interpretations. Skip logic was used within these surveys. The “relevant” column contains all skip logic related to that variable, and references other previously mentioned variables.

The choices tab relates to all multiple or single choice options available to the respondents. Again, these may have been numerically coded, but presented to the respondent in one of two languages. All choice labels can be found in this tab. The Settings tab states the form title, and also allows you to set the default language.

Within each country we had slight variations in the questions asked to ensure they were contextually relevant (except where we were following international indexes- see folder 1.iv SOS\_ Survey\_Design\_References.docx). Container based sanitation users also were asked different questions compared to non users. Therefore within each country we have multiple user types, and therefore we also had separate surveys which can be found in each of the country folders: CBS or NU (non user). See the table below for a summary of how many ODK forms can be found in each folder.

|  |  |
| --- | --- |
| **Forms** | **Number of versions**  |
| Demographics-BASELINE | 1 |
| Demographics- WEEKLY | 80 |
| Demographics- MONTHLY | 14 |
| Election | 80 |
| Electricity | 80 |
| Household (Poverty) | 4 |
| Journey\_task\_cbs\_collection | 1 |
| Journey\_task\_community\_toilet | 1 |
| Journey\_task\_truck\_road | 1 |
| Sanitation\_access\_cbs\_users | 80 |
| Sanitation\_access\_non\_users | 80 |
| Shocks | 80 |
| SNA\_CBS\_users baseline | 1 |
| SNA quarterly | 4 |
| Wash\_monthly | 14 |
| Wash\_once | 1 |
| Wash\_quarterly | 4 |
| Wash\_weekly | 80 |
| Wellbeing | 80 |

* + - 1. Kenya > English & Swahili
			2. Peru > English & Spanish
			3. South Africa > English & Xhosa

**Creator contact details**

Dr Amy Lewis

e-mail: amy.lewis@bangor.ac.uk

Professor Simon Willcock

e-mail: s.willcock@bangor.ac.uk

Dr Alison Parker

e-mail: a.parker@cranfield.ac.uk