

SurveyLab

A virtual environment for teaching survey methods and real-world data analysis

User Manual

Version 0.9, October 2022



Content

- 1. Introduction 2
- 2. Access 2
- 3. Conducting a virtual survey 6
 - 3.1. Tutorial 6
 - 3.2. Explore: get to know the environment 7
 - 3.3. Design: design the data collection tool 8
 - 3.4. Sample: define the sample strategy 10
 - 3.5. Survey: collect and download data 11
- 4. Technical Details 15
- 5. Credits 15

1. Introduction

Engaging students in real-world data production is important for teaching survey methods and data analysis. Active involvement in designing sampling strategies and data collection allows students to experience difficulties and costs of obtaining quality data, and the often neglected importance of proper data management. Seeing data together within the context they originate from allows for a deeper understanding of the strengths and limitations of the different analysis techniques. As such, the student gains a better appreciation of the strong interdependency between the various phases of a research project, from the definition of the sampling strategy and sample size calculation until the interpretation of the analysis results.

However, in practice, students have limited possibilities in this regard. The real-life research projects need a complex and time-consuming preliminary phase (including ethical approval or, in any case, adequate consideration for safety and health issues for the students/investigators themselves). The project often requires expenses that need to be funded, and when involving human subjects, there is a requirement for strong supervision by experienced researchers. All these aspects are hardly compatible with the short time frame and limited resources of a semester (or even annual) teaching module.

The *SurveyLab* is conceived as a tool to integrate and complement theoretical teaching of survey methods and traditional student assignment work, promote active learning and facilitate the development of practical competencies in designing and conducting population surveys. It consists of a virtual environment including a simulated human population of about 40000 individuals distributed across the 5 Regions and 21 Towns of a hypothetical country. The simulated population is not directly representative of any real country. However, the distributions of the variables (demographic, socioeconomic, and health-related) and their relationships are realistic and derived from the analysis of Demographic and Health Surveys conducted in low- and middle-income countries and the epidemiological literature regarding relationships between risk factors and diseases.

The *SurveyLab* environment is remotely accessible through a web interface. It allows users to design their survey projects, sample the population according to their plan, and collect the data of interest that they can then analyse and interpret.

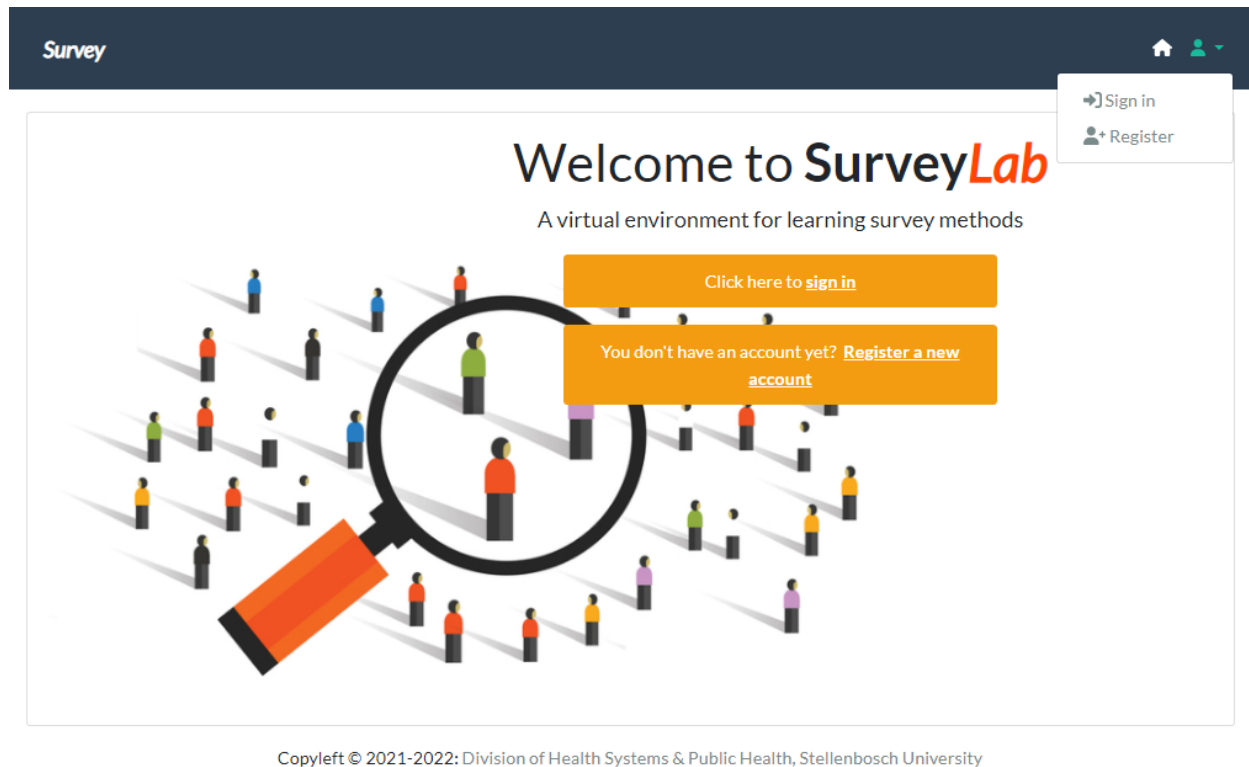
This **draft manual** provides instructions on how to access and use the *SurveyLab* for **testing and evaluation purposes**. As the environment is still in the evaluation phase, **the final version of the manual might be different**.

2. Access

The *SurveyLab* is freely accessible at the URL <https://surveylab.sun.ac.za>, with any modern internet browser supporting html5 and javascript. There are no technical limitations for access from mobile devices, and the interface will adapt to any screen size. However,

activities such as sampling, which requires exploring maps of the environment, are best conducted on large screens, and a minimum resolution of 1024x768 pixels is recommended.

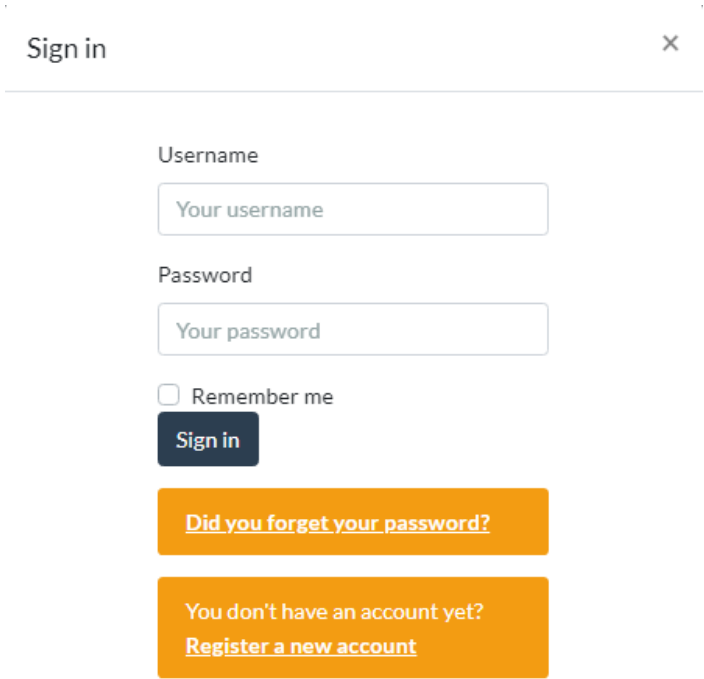
On the home page, you will find the option to **sign in** (if you already have an account) or **register**.



Clicking on **register** will open a registration page.

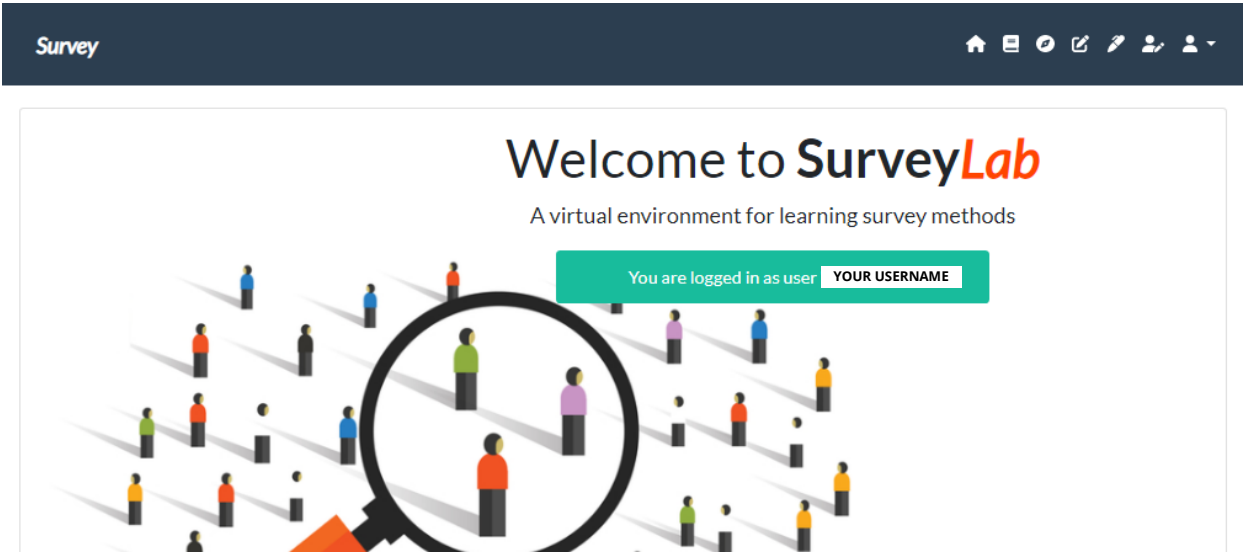
After filling out the form and clicking on the submission button, you will be added to the user list, and an email will be sent to verify your address. You can click on the link provided in the email or copy the URL in your browser address bar to validate the registration and gain full access to the system.

After the registration, you can use your username and password to sign in to the system. If you click on **sign-in** on the home page, you will be redirected to a form where you can type in your credentials:

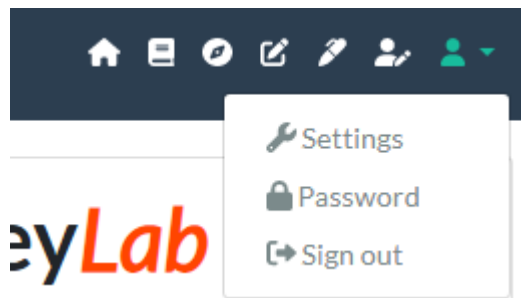


The image shows a 'Sign in' form with a title bar containing 'Sign in' and a close button 'x'. The form includes a 'Username' field with the placeholder text 'Your username', a 'Password' field with the placeholder text 'Your password', and a checkbox labeled 'Remember me'. Below these fields is a dark blue 'Sign in' button. At the bottom of the form, there are two orange buttons: one with the text 'Did you forget your password?' and another with 'You don't have an account yet? Register a new account'.

On the new home page that will be presented after signing in you will find confirmation of your username and a new set of menu items



The last menu item on the right allows you to manage your account.



Clicking in the **Settings** menu will show a form to change the username and email address.

A screenshot of the 'User settings for [YOUR USERNAME]' form. The form is displayed within a white container with a dark blue header bar at the top. The header bar contains the word 'Survey' on the left and navigation icons on the right. The form title is 'User settings for [YOUR USERNAME]'. Below the title, there are three input fields: 'First Name' with the value 'Annibale', 'Last Name' with the value 'Cois', and 'Email' with the value 'annibalecois@gmail.com'. Each input field has a green vertical bar on the left side. At the bottom of the form, there is a dark blue 'Save' button.

The **Password** menu allows you to change your password.

A screenshot of the 'Password for [annibale_1]' form. The form is displayed within a white container with a dark blue header bar at the top. The header bar contains the word 'Survey' on the left and navigation icons on the right. The form title is 'Password for [annibale_1]'. Below the title, there are three input fields: 'Current password', 'New password', and 'New password confirmation'. Each input field has a red vertical bar on the left side. At the bottom of the form, there is a dark blue 'Save' button.

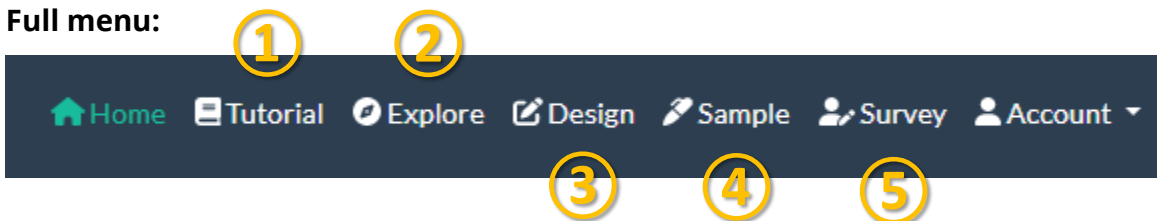
3. Conducting a virtual survey

The *SurveyLab* main menu is shown below in its condensed (small screens) and full format:

Condensed menu:



Full menu:



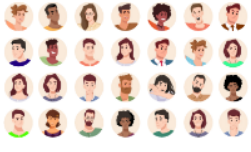



Items 1 (**Tutorial**) and 2 (**Explore**) link to web pages which provide information on the survey process and the environment. The process of conducting a virtual survey on the *SurveyLab* involves – similarly to a real-life survey – three steps: (1) designing the data collection tool; (2) defining a sampling strategy and (3) collecting data. These steps are implemented through items 3 (**Design**), 4 (**Sample**) and 5 (**Survey**) of the menu.

Below is a description of the specific features of each of these *SurveyLab* sections.

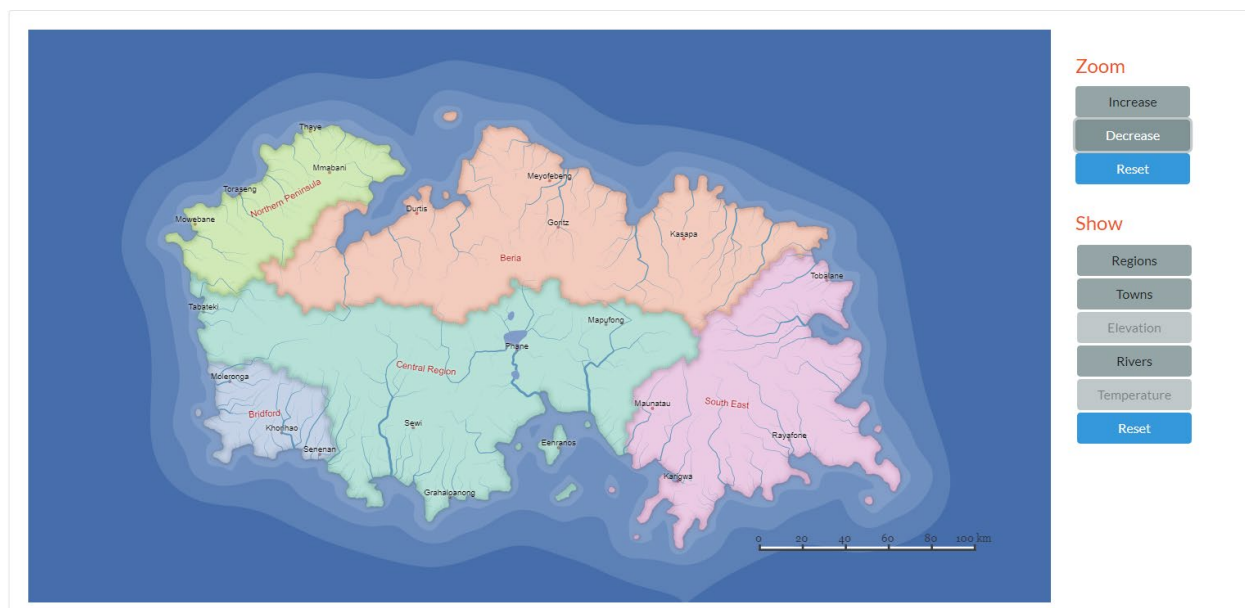
3.1. Tutorial

The **Tutorial** page includes some general information about the *SurveyLab* and, at the bottom, a link to four video tutorials, which will present in a more direct form the content of this manual. **Note that the videos are currently not implemented, as they will be finalised after the conclusion of the evaluation phase of the system you are contributing to.**

 <p>Explore</p> <p>Visualize a map of the virtual world, get information about towns, neighborhood and households.</p> <p>Video tutorial</p>	 <p>Design</p> <p>Plan your study and design your data collection tool.</p> <p>Video tutorial</p>	 <p>Sample</p> <p>Sample your population.</p> <p>Video tutorial</p>	 <p>Survey</p> <p>Collect data from your sample.</p> <p>Video tutorial</p>
--	---	--	--

3.2. Explore: get to know the environment

The **Explore** page presents a dynamic map of the *SurveyLab* world.



The group of buttons on the top right allows you to **increase/decrease** the map's scale. The interface also supports mouse operations and touch screens, so scrolling and zooming of the map can be also obtained using these devices. The **"reset"** button restores the initial settings. The groups of buttons on the bottom right allow you to show/hide different information types, **region names**, **town names**, **elevation**, **rivers and lakes**, and **average annual temperatures**. The **"reset"** button restores the initial settings.

Clicking on a **region name** visualises a pop-up window with information on the region, including sociodemographic indicators and the list of the various settlements (towns).

Region

Northern Peninsula

- Population: 7182
- Capital: Mmabani
- Average Wealth: 22.6
- Gini Index: 31.6
- Environmental Offices: 1
- Clinics: 5
- Hospitals: 1

Towns

Mmabani

- Population: 2415
- Fraction: 33.6%

Mosewane

Clicking on a **town name** opens a new page relative to the chosen town:



The buttons on the right-hand side function as on the previous page and allow for increasing/decreasing the map size and visualising/hiding different types of information.

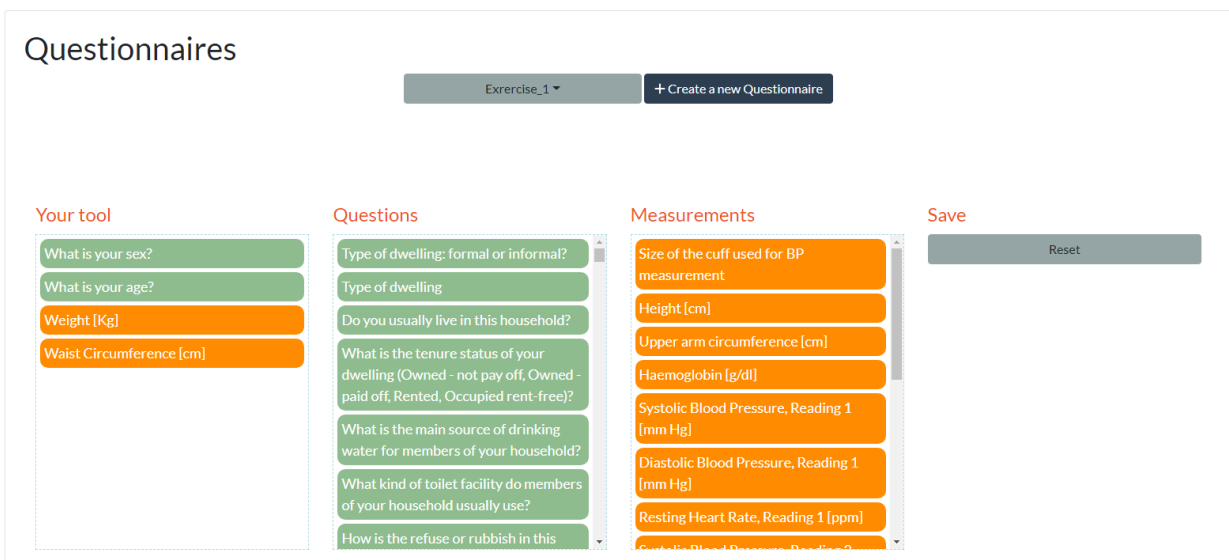
Clicking on the symbols below (**Public Buildings**) visualises pop-ups windows providing information on the Town (**Municipal Offices**, one per town), on the **Clinics/hospitals** (none, one or more per town) and about the environment (**Environmental Health Offices**, present only in the capital town of each region)



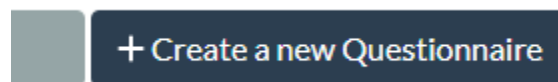
The information provides a basis for planning the survey.

3.3. Design: design the data collection tool

The **Design** page allows for designing a data collection tool for the survey. The tools can include both questionnaire items and measurements and can be built by dragging the relevant questions/measurements from the lists on the right and dropping them on the left section (**Your tool**).



To create a new list, click the button on the top right-hand corner of the screen, as illustrated below:



By selecting the tab **+create a new questionnaire**, you will open a new page where you can name your questionnaire (**Description**), **activate** it, and **save** it for future use.

Create or edit a Questionnaire

Description

Example Questionnaire

Activated



Cancel

Save

Before saving, ensure that the flag **Activated** is selected, otherwise you won't be able to use your questionnaire.

3.4. Sample: define the sample strategy.

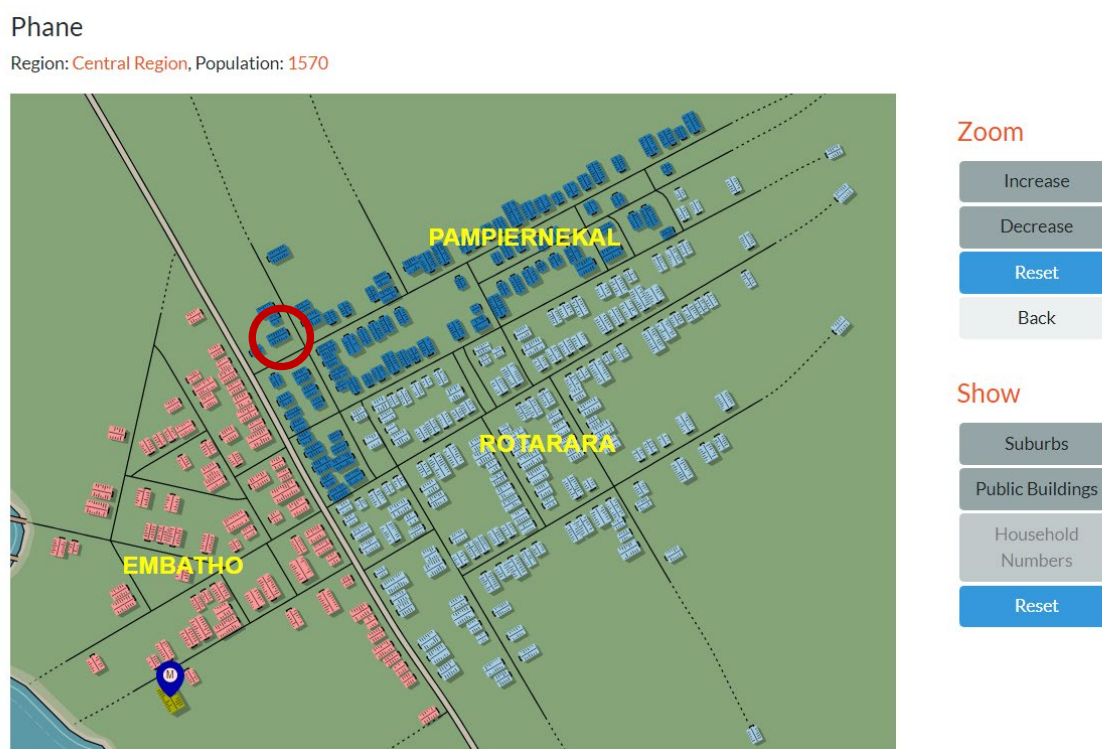
The **Sample** page allows for selecting the sample to be surveyed. Note that the *SurveyLab* does not provide automatic means for sampling selection, as it is designed as a learning tool where the users/students design their sampling strategy.

The sample selection includes different steps:


1. Selection of **towns**;
2. Selection of **households** within a town;
3. Selection of **individuals** within a household.

The sample page looks and works similarly to the explore page, but, in this case, clicking on a town name leads to a page where it is possible to select households (and individuals within households) to be part of the sample.

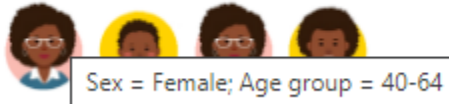
The households are selected by clicking on a dwelling: the dwelling change colour, the whole household is added to the sample, and the members are listed at the bottom of the page.




For example, selecting the dwelling circled in the figure above makes this list appear below the map:

Households • H_18_263 

The list includes the household's unique identifier and a symbolic picture of its members. Hovering with the mouse on a picture makes visible the basic demographic of the member:

Households • H_18_263 

By default, all members of the household are added to the sample, but it is possible to deselect any individual by clicking on the picture. For example, in the figure below the second individual of household H_18_263 is deselected and will not be included in the sample:


Households • H_18_263 

This procedure can be repeated many times to select multiple households in a town and to add multiple towns to the overall sample.


3.5. Survey: collect and download data

When the two preliminary phases of designing the data collection tool and selecting the sample have been completed, the **Survey** page allows you to start the (virtual) data collection.


Overview



Data collection tool
Test 1 ▾
Number of questions: 4
Number of measurements: 2
Administration time: 12




Sample
Number of towns: 3
Number of households: 11
Number of individuals: 41



Budget
Set up cost: 9
Collection cost: 0.7
Total cost: 9.7
Unit cost: 0.24

Survey

Choose your **data collection tool** among those you have previously saved, and click on the 'survey' button. When the system has completed the survey (the time needed depends on the sample size and the length of the questionnaire), you can **download the data** by clicking the "download" button



Households visited: 0%

Individuals interviewed: 0%

Download data

The top part of the page includes (1) a summary of the data collection tool; (2) the size of the selected sample at the household and individual level, and (3) the expected cost of the survey. For example, in the simplified case illustrated in the figure above, the summary indicates that:

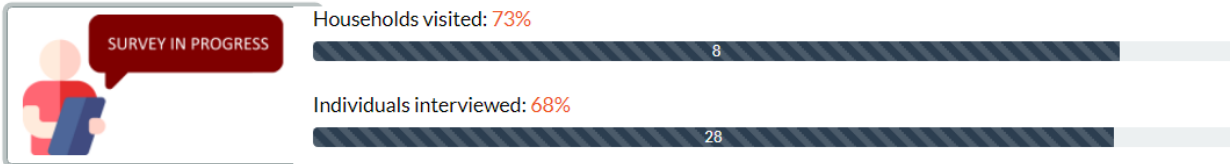
- 1. The **data collection tool** consists of 4 questions plus two direct measurements for an estimated average administration time of 12 minutes per participant;
- 2. The **sample** is spread across 3 towns and includes a total of 11 households and 41 individuals;
- 3. The total **cost of the survey** is 9.7 cost units (imaginary currency of the country), of which 9 are due to setup (admin, transport, training of fieldworkers, ...) and 0.7 for the actual collection. The cost per sampled individual is estimated at 0.24 cost units.

The system estimates administration times and costs using an algorithm that attributes different weights to different types of items in the data collection tool (e.g. it attributes longer administration times and costs to anthropometric measurements compared to questions). The calculations takes into account increases in setup costs when a survey spans across multiple towns and the costs within each town based on the total distance a fieldworker has to walk/drive to visit each selected household.

The large **button** on the left allows you to administer the survey.

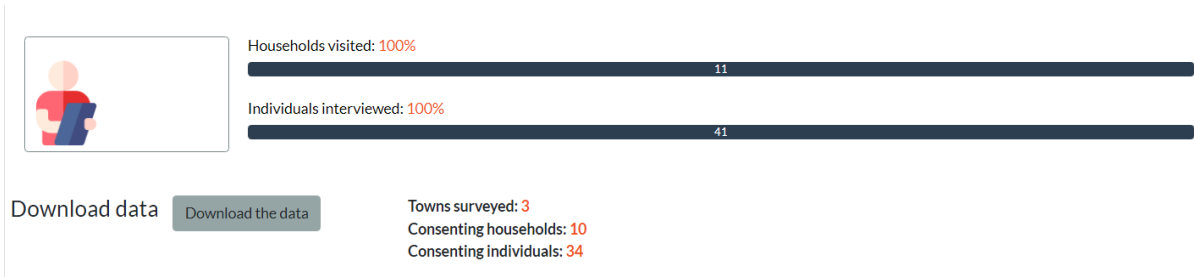


When pressed, the data collection starts, and the **progress bars** shows the percentage completed.



The time needed to complete the survey depends on the length/type of data collection tool and the sample characteristics, which gives an indication of the total 'effort' to complete the data collection. However, it must be noted that the actual absolute speed of completion is also affected by the user's computer; therefore, administration times cannot be compared across *SurveyLab* users.

When the data collection is completed, a summary of the results of the survey is shown, and the **download** button becomes active:



Note that the **number of households and individuals from which data are collected** (respondents) **usually does not match the sample size**, as in this case.

This is because the virtual households and individuals in the *SurveyLab* can refuse to participate in the survey (or withdraw their consent after an initial acceptance), simulating the behaviour of human respondents. The algorithm attributes the probability of refusal according to (1) individual and contextual sociodemographic characteristics and (2) the length of the data collection tool and type of items. The actual consent/refusal is calculated every time the user requests data collection. It is random (based on the probabilities described above), resulting in patterns of consent/refusals that are generally different for each data collection, even with the same sample.

Clicking on the **download** button creates a **comma-separated value (.csv)** file, which is downloaded to the user's machine. The file is compatible with Microsoft™ Excel®.

The format of the file is shown in the figure:

①	②	③	④	⑤						
IID	HID	HR	IR	DWELLING_TYPE	HEIGHT	WEIGHT	ENERGY_COOK	SEX	AGE	
H_4_308_1	H_4_308	Consent	Consent	Formal	168.2	80.3	Electricity from mains	Male	25	
H_4_333_3	H_4_333	Consent	Consent	Formal	176.1	88.5	Electricity from mains	Male	32	
H_4_327_4	H_4_327	Refusal		null	null	null	null	null	null	
H_4_327_5	H_4_327	Refusal		null	null	null	null	null	null	
H_4_327_6	H_4_327	Refusal		null	null	null	null	null	null	
H_4_331_7	H_4_331	Consent	Consent	Formal	159	69.4	null	Male	null	
H_4_331_8	H_4_331	Consent	Refusal	null	null	null	null	null	null	
H_4_333_1	H_4_333	Consent	Consent	null	165.1	64.7	Electricity from mains	Female	78	
H_4_333_2	H_4_333	Consent	Consent	null	144.6	93.1	Electricity from mains	Female	25	
H_4_331_2	H_4_331	Consent	Consent	Formal	158	82.7	Wood	Female	72	
H_4_325_5	H_4_325	Consent	Consent	Formal	null	null	Wood	null	5	
H_4_325_6	H_4_325	Consent	Consent	Formal	null	null	Wood	Male	10	
H_4_327_1	H_4_327	Refusal		null	null	null	null	null	null	
H_4_327_2	H_4_327	Refusal		null	null	null	null	null	null	
H_4_327_3	H_4_327	Refusal		null	null	null	null	null	null	
H_4_331_3	H_4_331	Consent	Consent	null	155.4	69.5	Wood	Female	46	
H_4_331_4	H_4_331	Consent	Consent	Formal	null	null	Wood	Female	3	
H_4_325_1	H_4_325	Consent	Consent	Formal	158.7	115.9	Wood	Female	55	
H_4_325_2	H_4_325	Consent	Consent	Formal	154.2	null	Wood	null	29	
H_4_325_3	H_4_325	Consent	Consent	Formal	null	null	Wood	null	14	
H_4_325_4	H_4_325	Consent	Consent	Formal	null	null	Wood	Male	10	

Column 1 (**IID**) contains the **individual unique identifier**, column 2 (**HID**) the **household unique identifier**, column 3 (**HR**) shows **consent/refusal at the household level**, and column 4 (**IR**) **consent/refusal at the individual level**. The remaining columns (5) match the questions/measurements in the data collection tool, where '**null**' indicates **missing data**.

In the **IIH** and **HID**, the first number is the town identifier, according to the following table:

Code	Town Name
1	Durtis
2	Meyofebeng
3	Goritz
4	Kasapa
5	Moleronga
6	Khorihao
7	Senenan
8	Mowebane
9	Toraseng
10	Mmabani
11	Thaye
12	Maunatau
13	Karigwa
14	Rayafone
15	Tobalane
16	Tabateki
17	Sewi
18	Phane
19	Mapufong
20	Grahaloanong
21	Eenranos

4. Technical Details

With the objective of long-term sustainability, portability on different servers (local or cloud-based) and ease of integration on various learning platforms, the system is completely developed using open-source, license-free technologies.

The graphical interface is coded using HTML5/Javascript and standard libraries. The backend server (which manages the virtual population and responds to requests for sampling and downloading of respondent characteristics) is implemented using R statistical language v. 4.0 (R Foundation for Statistical Computing Platform, Vienna, Austria). It exposes a standard interface compliant with the REST API protocol, allowing for a complete uncoupling with the interface (and the concurrent use of multiple interfaces).

5. Credits

The *SurveyLab* is a product of the **Division of Health Systems and Public Health** at Stellenbosch University. The initial development has been funded through the **Fund for Innovation and Research into Teaching and Learning** (FIRTL) of the University (<http://www.sun.ac.za/english/learning-teaching/ctl/t-l-awards-and-grants/firtl>).

Dr Annibale Cois (acois@sun.ac.za) conceptualised the project, led the development, generated the virtual population, and wrote the backend R code and the REST API interface. **A/prof Lungiswa Nkonki** (lnkonki@sun.ac.za) collaborated in defining the characteristics of the virtual population and designing and implementing the evaluation process. The **Research and Social Impact Systems unit** at Stellenbosch University supported the development process, contributed to the engineering of the original idea **and coordinated the interaction with the IT department**. The **IT Institutional Software Solutions unit** at Stellenbosch University coded and deployed the web interface, supported the servers' cloud deployment, and ensured maintenance.