

Title:

Data generating project on informal sector activities in urban areas in South Africa

Institutional Partners

Southern Africa Labour and Development Research Unit (SALDRU), University of Cape Town

Sustainable Livelihoods Foundation (SLF)

Participants:

Andrew Charman, Dr. Charman is a social scientist and development practitioner at the Sustainable Livelihoods Foundation.

Leif Petersen, Mr. Petersen is an eco-sociologist and development practitioner at the Sustainable Livelihoods Foundation.

Vimal Ranchhod, Dr. Ranchhod is an Associate Prof. in SALDRU at the University of Cape Town. He will be the co-ordinator of the activities undertaken for this project.

Description of Project:

The overall objective of the project is to generate, from a number of censuses, a consistent dataset that measures the amount of informal sector activity occurring in eight locations in urban townships in South Africa. This includes an enumeration of the number of informal sector enterprises, as well as some information about their activities, volume of trade, barriers to growth and the amount of employment that they generate. The methodology used involves area-censuses, whereby small geographic areas were selected and all activities occurring within them were enumerated. These censuses have already been conducted and captured in Excel by the Sustainable Livelihoods Foundation. We are now seeking funding to process this data into a consistent and usable dataset.

Relevance to REDI3x3:

While the work planned for this project is purely focused on data generation, the subsequent research that the data will enable will fall clearly within the stated objectives of REDI3x3 in general and ISEP in particular. We are thus applying for funding for a *special project* under the *Employment/Unemployment* focus area.

The specific gaps that this dataset could help to fill include:

- Labour demand and firm behaviour in the formal and informal sectors, urban and rural, in various sectors of the economy.
- Reforming the measurement of unemployment, underemployment, employment and related livelihood states.

While the gaps above are very broad in scope, we believe that the preparation of this dataset could make a modest contribution to these gaps as follows:

- i) Improving our understanding and characterisation of firm structure amongst micro-enterprises
- ii) Understanding aspects of employment within micro-enterprises
- iii) Understanding the spatial dynamics of firm distribution
- iv) Understanding livelihood streams through business activities within the township informal economy.

Deliverables

The final output of this project is a dataset ready for public release, drawn from the eight area-censuses that have been conducted by the Sustainable Livelihoods Foundation. This dataset will be made available in Stata, Excel, SAS and SPSS format. The dataset will be called the SLF Micro-Enterprise Study. The decision to release the dataset publicly within the context of REDI3x3/DataFirst rests with the Director of REDI3x3.

Along with the dataset, a set of documentation will be released for third party users. This will include a printable version of the questionnaires and a User Guide & Metadata document that explains how variables are coded from the questionnaire. It will also include notes on post-survey derivation of any of the variables included in the release.

It is agreed that REDI3x3, having invested in the processing of the existing survey data to enable a public release to researchers working on key research questions regarding the informal sector within REDI3x3 and ISEP (but not excluding other researchers from SLF or elsewhere), will be consulted in further SLF area-censuses insofar as survey, sampling and questionnaire design etc. is concerned, and that subsequently-gathered survey data will continue to be made available to in a similar format and manner as the current dataset (subject to funding being available for the processing).

Structure of the data

- The data comes from a census of a particular area.
- Eight areas have thus far been surveyed.
- Within each area, all enterprises were enumerated.
- If the activities of an enterprise fell within a particular industry, for example Spaza Shops, the owner/employee were then asked a set of questions based on the “Spaza Shop” module.
- Thus, only enterprises that fell within the scope of a module were surveyed.
- Nonetheless, the modules were designed to include the most common forms of activities.
- A questionnaire consists of various modules.

- With time, the modules included in the questionnaire changed, and the questions asked within a module changed. Either of these would result in a new version of the questionnaire.
- In total we have five different versions of the questionnaire.

The variations between areas, modules and questionnaires are shown in Table 1 below. We observe that the module on Hair Salons were only included from version 3 of the questionnaire onwards, and was thus administered in only 4 of the Areas. Moreover, there may potentially be some differences in the questions asked of Hair Salons in some areas, although the same questions will have been asked in Ivory Park and Tembisa.

Limitations

The data has some information that is considered to be sensitive by the original surveyors. These generally relate to the ability to identify the respondents. We thus require that the public release will be stripped of all variables that could be used to identify respondents. This includes their names, telephone numbers, photographs of the enterprises, the precise GPS co-ordinates of the enterprises, physical location of the enterprise and the name of the enterprise.

These are the variables which we consider to be problematic as this point, but there may be others that we realise are problematic with regards to the anonymity of the respondents at a later stage. We thus further reserve the right to either remove, or sufficiently aggregate, any other such variables at a later stage. This will be done in consultation with the REDI3x3 research co-ordinator.

At the same time, we recognize that one of the most valuable components of this dataset is the geographic location information. We thus propose that the GPS data be released with an accuracy level of up to 50 metres from the actual point measured by the interviewers. This is a compromise that will facilitate usage of the data but not compromise the identity of the respondents.

There is also the possibility for researchers to apply to have access to the confidential version of the data. This would be housed in Data First's secure data enclave facility. Applications will be considered on a case-by-case basis. The decision to grant access will be made by SLF in consultation with REDI3x3 initially, for a period of 24 months, and thereafter such authority will be ceded to Data First. Relevant criteria for making such decisions include whether any interviewees will be exposed to any risk, the importance of the research question, the ability to obtain the required information from another dataset, the feasibility of the proposed research with and without the confidential data, and the willingness of the researcher to collaborate with the SLF-REDI3x3 research team.

Table 1: Areas, modules and version of questionnaire

Module	Area							
	Delft South	Vrygrond	Sweet Home Farm	Browns Farm	Ivory Park	Tembisa	Imizamo Yethu	Kwa Mashu
Spaza - price	1	1	1	2	3	3	4	5
Spaza - violence	1	1	1	2				
Liquor Retailers	1	1	1	2	3	3	4	5
Educare	1	1	1	2	3	3	4	5
Traditional healers	1	1	1	2	3	3	4	5
Hair Salons					3	3	4	5
Street Traders					3	3	4	5
Micro-manufacturing								5

Notes:

1. The number in the cell designates which version of the questionnaire the module comes from.
2. There are five versions of the questionnaire.
3. Not all modules were administered in all areas.
4. An empty cell signifies that this module was not administered in this area.

Activities

The specific activities planned for this project include:

- Coding of variables for missing values
- Standardize the variables across modules within questionnaires
- Standardize the variables across questionnaires
- Coding of qualitative responses
- Coding of photographic data
- Writing a User guide/Meta-data document
- Formatting and construction of a printable document for each version of the questionnaire

Timelines

We expect that the process will take three months to complete, from the time that the funding is approved.

Budget

Our overall budget for this project is 70 460 rand. This is decomposed into its various sub-components in Table 2 below. The RA is priced at the level of an honours graduate, Liedeman at the rate of a masters graduate, Charman and Petersen at the level of a lecturer, and Ranchhod at the rate of an Associate Professor. The SLF component (Charman/Petersen and Liedeman) will be paid to the Sustainable Livelihood Foundation as a lump sum upon completion of this project.

The responsibilities of each team member in this project are as follows:

- Ranchhod: Overall management and coordination responsibility, overall design of the final dataset, converting the data from Excel into Stata format and thereafter into SAS, SPSS and Excel format, monitoring and reporting on data-base scope and scale, development of a User guide and Metadata document.
- Research Assistant: Data entry and data recoding; harmonisation of data-bases into single data base, assist Ranchhod with the Stata coding required and with compilation of the User guide and Metadata document.
- Charman / Petersen: Survey design, supervision of data entry process, translation of qualitative data to quantitative data and description of User Guide.
- Liedeman – data proofing and verification, data entry, translation of qualitative to quantitative.

Table 2: Budget

Personnel costs	Hours	Hourly rate	Sub-total
Research Assistant (tbd)	176	R 120	R21120
Charman / Petersen (combined)	40	R 220	R 8800
Ranchhod	80	R 319	R25520
Liedeman (SLF assistant)	80	R 150	R12000
Software: Stata license			R 3020
Grand Total	376		R70460