Data Infrastructure Proposal

Andrew Kerr, DataFirst

Creating a Harmonised Survey of Employers and Self-employed dataset

The aim of this work is to create a harmonised version of the four Surveys of Employers and the Self-Employed and make the data available through the DataFirst website.

The Survey of Employers and Self-Employed (SESE) is a survey of firms not registered for VAT that is undertaken by Statistics South Africa every 4 years. There are currently 3 surveys in the public domain (2001, 2005, 2009) whilst the 2013 survey has been done but the data is not yet in the public domain.

The SESEs are one of the few sources of firm data available in South Africa. Questions are asked about the owner and the firm. It is possible to link the owner to his or her household, since the sample frame was the Labour Force Survey conducted at the same time as the SESE. It is envisaged that the harmonised version of the SESE datasets will contain information about the owner's household from the (Q)LFS in addition to the data from the actual SESE.

The main activity for this project will be to clean and code the three SESEs that are currently publicly available. Updating the data when the forth SESE is released and cleaning and coding this new dataset is the other main activity of the project and will be a second part of the project. There will also be documentation produced, to be released with the data, explaining how the data was cleaned and coded, how it can be used and what is included in the data. It is anticipated that these activities will require fifteen days of the Senior Research Officer's time.

The creation of metadata and dissemination of the data on the DataFirst data portal is also required. It is anticipated that this will require one day of the DataFirst Manager's time.

The final requirement is fielding queries on the dataset, which is handled by the DataFirst technical officers. It is anticipated that this will require one day of technical officer's time.

Expenses

	Cost
Cleaning and Coding the SESE	
Senior Research Officer (15 days @ R4000)	R 60 000
Metadata creation and data dissemination	
DataFirst manager (1 day @ R6000)	R 6 000
Data queries	
Technical Officers (1 day @ R 2000)	R 2 000
Total	R 68 000

The expected completion date for the first part of the project, making the first three SESEs available on the DataFirst portal, is 31 March 2014. The second part involves updating the dataset using the 2013 SESE. If the data is released in the first half of 2013 then the completion date would be 30 September 2014.