DOMESTIC LOAD RESEARCH DATA COLLECTION USING OPEN DATA KIT

Schalk Heunis, Marcus Dekenah, Lloyd Setlhogo
Enerweb – Demand Intelligence, Eskom Holdings SOC Ltd

ABSTRACT

Open Data Kit (ODK) is a free and open source set of tools that simplifies the collection of survey based data. During the operations of the NRS Domestic Load Research programme, a number of survey based tasks are executed and is mostly paper based. ODK was piloted for the collection of site audit information for the NRS load research project and the tools used and pilot results obtained are described in this paper.

ODK comprises three stages – build, collect and aggregate. During build the survey form is defined in an open format. ODK collect uses the form definition, to dynamically create an interface for collecting data using the mobile devices using the Android operating system. During aggregation the results are exported for use in spreadsheet programmes and GIS applications.

The site audit survey process was implemented using ODK and the data collection process was successfully completed for all active 2013 sites. It is currently being deployed for other survey processes of the NRS Domestic Load Research programme.

This toolset is recommended for all survey based data collection and could add value during energy audits, customer interviews and device installations

Keywords— **Load research, survey based data collection, mobile device application, free and open source software**

1. INTRODUCTION

The NRS domestic load research programme [1] collects load readings and statements of their key socio-demographic indicators from customers in an effort to satisfy a national sample design [2]. The sample design is based on household income, time since electrification and climatic region and the aim of the programme is to research the load usage of a representative sample of the national population in order to produce guidelines for planning and design activities.

Apart from the 5 minute electrical usage (in W, VA, A and V) per customer, quantitative data is collected during the planning and operational phases of the load research activities of the NRS load research programme and linked to the customer, for example:

- Site survey – the aim being to identify potential load research sites and validate that the customers socio-demographic details correspond with those as identified by the sample design
- Site referencing – the aim being to validate the installation, connection and metering details of customers.
- Socio demographic survey – the aim being to capture a socio-demographic description of each household and linking that back to the electrical load usage.

This quantitative data is currently collected using a paper based process, then captured electronically, validated and linked. This paper describes the use of Open Data Kit (ODK) as a replacement of the paper based methods for domestic load research data collection.

The following figure shows the as-is and to-be data collection processes:

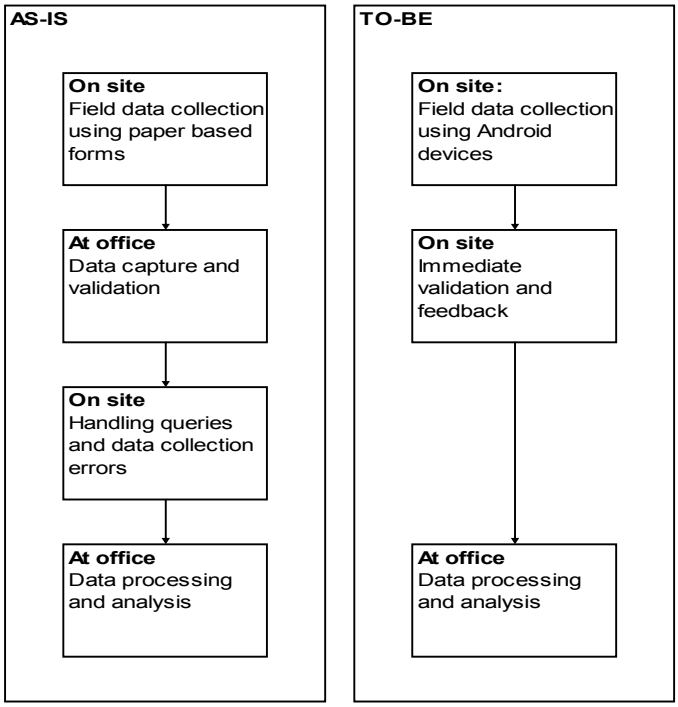


Figure 1: As-is (paper-based) and To-be (ODK based) data collection processes

2. OVERVIEW OF OPEN DATA KIT

Open Data Kit (ODK) "... is a free and open-source set of tools which help organizations author, field, and manage mobile data collection solutions." [3]

From [3]: *ODK provides an out-of-the-box solution for users to:*

- *Build a data collection form or survey*
- *Collect the data on a mobile device and send it to a server; and*
- *Aggregate the collected data on a server and extract it in useful formats.*

ODK was initiated as part of a Google (google.org) sponsored sabbatical project and the core developers are researchers at the University of Washington's Department of Computer science and engineering and associated members. Their purpose is to improve the lives of under-served populations around the world.

The project is mature and is currently at version 1.4 with active development on version 2.0

ODK simplifies the collection of data by using

- a standardized form definition (XForms) to define what information is to be collected,
- publishing and reading the form definition from a web based server using the "Collect" Android application
- Using the form definition on the Android device to collect data fields, applying what data validation rules are defined,
- Sending the collected form data to the server for analysis
- Post-processing the form data on the server and downloading the data for further analysis.

These steps are handled by three components and each is described in more detail in the following sections:

2.1 BUILD

During the build phase, templates (or forms) are defined for completion. Each template consists of fields which can have any one of a number of predefined data types, for example:

- Text field – any text, validation rules can be specified
- Image – either take a photo using the Android device or upload an existing image
- Geo-point – capture the current location using the Android device's GPS functionality

The fields can be ordered into groups which allow cascading questions, group repeats etc.

A number of different tools exist for form definition, for example:

- XLS2XForm
- Kobo
- PurcForms

These tools were evaluated and XLS2XForm was preferred for the definition of the Domestic Load Research project.

Once the template is complete it can be uploaded to a cloud based web server or self-hosted service and data collection can start.

2.2 COLLECT

The form definitions that have been uploaded, enables data collection to begin. ODK Collect is an Android application that reads the XForm definitions and dynamically produces an user interface for data capture.
For example the following is a screenshot of a simple form defined to capture site-survey information:

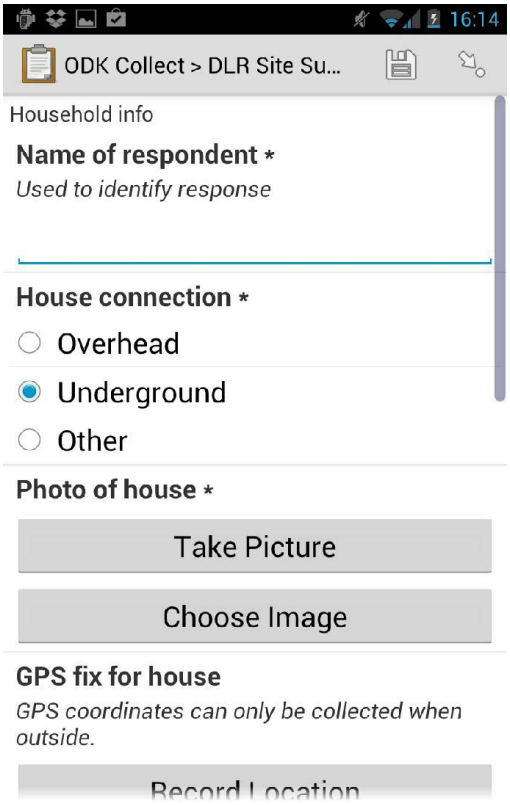


Figure 2: Screenshot of ODK Collect, configured with domestic load research site-survey questionnaire.

During the template definition phase, the designer can specify how many fields should appear on a screen – for example if a device with a small screen is used, each question may be placed on a separate screen.

A partially completed form can be saved for later retrieval and completion. Once the surveyor is satisfied with the contents, the completed form can be uploaded to a server or copied to a team-leader's laptop (if no internet service is available)

2.3 AGGREGATE

ODK Aggregate is a web service that can be deployed as cloud service or self-hosted service. It is used to:

- Publish form templates
- Upload completed forms
- Download results for analysis
- Publish results online for instantaneous reporting

ODK Aggregate is password protected, with role based user management, where typical roles include collection, deletion, publishing, publishing templates etc.

Data may be previewed on-line and downloaded in a variety of formats including

- KML – for viewing in Google Earth (including photos that were taken)
- CSV – analysis in a spreadsheet application
- XML – for use in downstream applications

3. CASE STUDY SITE REFERENCE INFORMATION

This section describes how ODK was applied in the NRS Domestic Load Research programme, lessons learnt and results achieved.

One of the operational tasks in the NRS Domestic Load Research programme is auditing the connections between loggers and households, in order to

- Verify the location of electrical usage data loggers, i.e. the GPS location of the pole and recorder identifier mounted on the pole
- Verify the households for which the electrical load data is recorded – the GPS location for the home, the electricity meter and the householders name is captured
- Verify which channel of the data recorder is connected with which house – this is done through either physically tracing the wire, or using a load test in the house.

This information is compared against installation data – in theory there should be no difference, but due to operational activities by the load research team and Eskom maintenance teams, differences have occurred.
Once the information has been verified and all queries resolved, the socio-demographic data collection team moves into the field to capture key indicators about the household during a 10-minute front door interview.
One challenge is to ensure that the socio-demographic team returns to the correct house. This is where the ODK solution adds immediate value – a photo of the house alongside a GPS location is captured and stored with the rest of the completed data. This increases the accuracy and repeatability of the entire data collection effort.

The following sections describe details of how each of the ODK steps was implemented:

3.1 BUILD

The following fields were defined using XLS2XForms:

- Recorder information: village name, recorder identifier, Eskom pole number, pole location (GPS fix), number of channels used
- Household information: channel number (for the specified recorder), photo of house (taking using the Android cameras), electricity meter number, house location (GPS fix), Eskom house number, address
- Link information: how was link established, i.e. physical trace or load test

For each recorder 2-3 households are connected and linked.
Specific validations were implemented on the following fields:

- Meter no – check that it is 11 numbers
- Recorder identifier – a pre-populated list ensures duplication and typographic errors

Further validations could be added, but this reduces flexibility and may force the data collector to enter invalid data in order to satisfy a rule. For example, by forcing a picture to be added – should the house owner object, then the data collector would not be able to complete the form, without taking a fake photo.

3.2 COLLECT

ODK Collect was loaded on a Android based tablet and two screens specified:

- Recorder level information
- Household level information

The latter group was repeated three times for each of the possible channels of the recorder.
The following figure shows a screenshot of ODK Collect, configured with domestic load research site audit questionnaire – showing the household information screen.

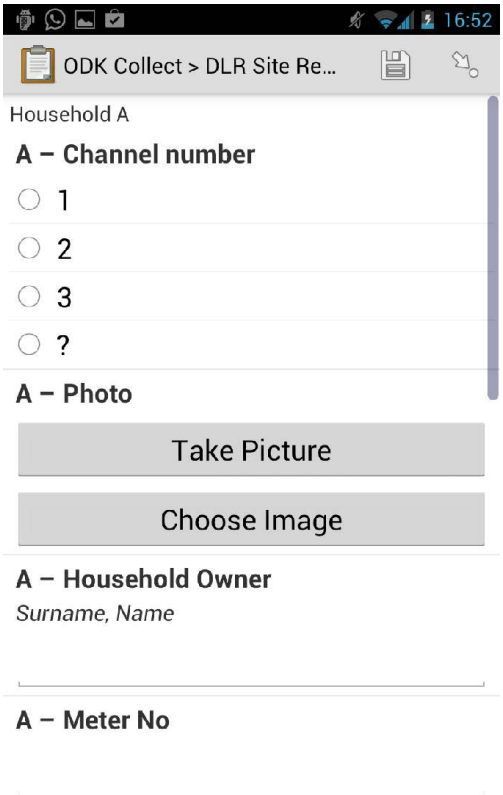


Figure 3: Screenshot of ODK Collect, configured with domestic load research site audit questionnaire – showing the household information screen

One of the differences between the Android based process and the paper based process is that the Android device has value beyond the data collected, i.e. the device could be stolen. Apart from the physical loss, all data not committed to the server would be lost.

In order to mitigate the risk of data loss, an application that continually syncs the content of the data collection folder to a second server was implemented. This background application is activated once anything in the ODK collect folder changes and uploads the data as soon as possible to an online back-up provider. Dropbox [7] was selected for use during the ODK pilot and DropSync [8] was selected for use on the device for continuous backup.

3.3 AGGREGATE

The completed forms were loaded to a cloud based aggregate service. The online publishing was not utilized and data was downloaded for analysis and publishing in a spreadsheet and Google Earth.

Since both recorder level and household level information was captured, the location and supporting information can be visualized separately.

The following screenshots were taken in Google Earth where the exported results are visualized.

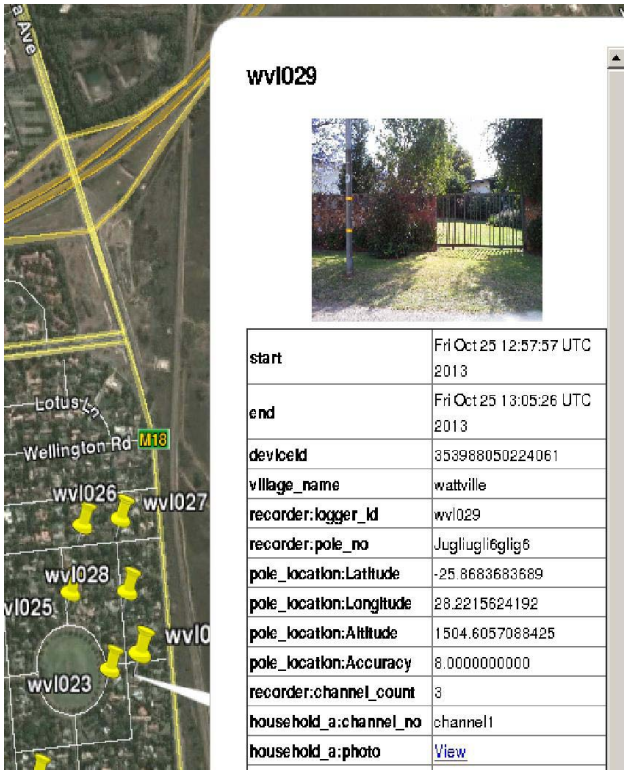


Figure 4: Screenshot of Google Earth showing recorder level information

Figure 5: Screenshot of Google Earth showing household level information

A version of Google Earth is available for Android and that means the information could be referenced on site by socio-demographic team in order to

- find household where surveys should be conducted
- plan their collection efforts, i.e. visualizing the spatial distribution of houses
- verify that the correct household is being interviewed
- correlate site audit data with similar responses, e.g. household head name.

4. CONCLUSION

An ODK data collection pilot for collecting domestic load research site audit data was successfully completed for all the 2013 load collection sites.

The collected data were used to verify data from the previous year and to assists in the identification of new sites. The geo-tagged photos returned are immensely valuable and allows office-based researchers to visualize the households investigated.

Using ODK, the cost of collecting high quality was significantly reduced, since the data verification and query phases were removed.

The ODK tool set has a learning curve and is steepest to grasp concepts behind XForms, i.e. how to properly define fields, validations, groups etc. XLS2Forms is definitely recommended for ease of use during the design phase.

Apart from simplifying and enriching the data collection process, additional benefits were realised, for example the socio-data collection team can readily use this data during that phase of the project.

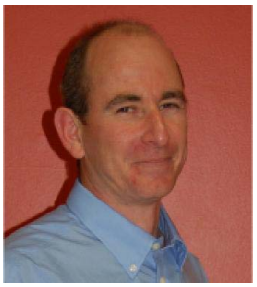
This work was funded by Eskom, and directed by the DSM, Energy Efficiency & Customer Research Steering Committee under Amal Khatri, (Activity Leader: Lloyd Setlhogo).

5. REFERENCES

[1] Dekenah, M; Setlhogo,L; "Domestic Load Research", Eskom Research and Innovation Department, Technology Strategy and Planning
[2] Heunis, S.Domestic Load Research Sample Design, Eskom, Jun 2009.
[3] Open Data Kit, http://opendatakit.org/about/
[4] XLS2XForm, http://opendatakit.org/use/xlsform/
[5] Kobo Toolbox, http://www.kobotoolbox.org/
[6] PurcForms, http://code.google.com/p/purcforms/
[7] Dropbox, https://www.dropbox.com/
[8] DropSync, https://play.google.com/store/apps/details?id=com.ttxapps.dropsync

7. AUTHORS

Principal Author: Schalk Heunis Holds a PhD in probabilistic modeling of power systems from the University of Stellenbosch (South Africa). He obtained his B.Eng and M.Eng degrees in Electrical Engineering from the University of Pretoria in 1994 and 1998.



In 2003 he joined Enerweb, where his responsibilities included load research, software architecting and system design. During this time, he headed the design team for the implementation of the Virtual Power Station for Eskom.

In 2011 he co-founded House4Hack with the purpose to promote innovation through open hardware and software. House4Hack has been involved and has supported winners of the first and third Gauteng innovation competition, Rorotika innovation to change the world andthe Gada prize for interim personal manufacturing.

His current area of interest is analytics for the internet of things.

Co-author: Marcus Dekenah holds a NHD ElecEng HC (Cape Technikon, 1983), M. Dip. Tech. Eng. (Cape Technikon, 1987), a BSc Eng (1st Cl Hons, UCT, 1991) and MBL II (Unisa, 1996).

Marcus Started his career at the Electricity department at Cape Town City Council involved in a wide range of distribution operations (1980-1988). From 1991- 1995 Marcus was executing contract research projects in HV testing, Energy management and Load research at Electric Power programme, Division of Energy Technology, CSIR.

In 1996 Marcus formed his own company and has since conducted contract load research (domestic, and non-domestic) both inside and outside South Africa with multi-disciplinary teams of specialists from industry, private, and the academic sector.

Marcus is currently assisting the Demand Intelligence Group of Enerweb with the execution of load research. Enerweb is a Southern African company focused on delivering solutions and services that enable utilities to manage demand flexibility.

Co-author: Lloyd Setlhogo Holds a ND ElecEng HC (Pretoria Technikon, 1997), B. Tech Power Eng (Unisa, Final Year).

Lloyd started his career at Technology Services International (TSI) a Division of Eskom Enterprises in 2000 as a Senior Technician. In 2005 Lloyd became the Load Research Project Leader. Lloyd was then promoted to Senior Advisor Engineering at Eskom Research, Testing and Development (2012).

Lloyd is currently managing the Load Research Project in Eskom Research (Demand Management Section) and has a wealth of experience in Eskom of 14 years.

Presenter: The paper is presented by Schalk Heunis