1.2 Open data kit for diagnostic crop production survey at landscape level in India

A. Rai[1], A. Ajay[1], S. Sharma[1], Soumen Pal, Alka Arora[2], Sudeep Marwaha, S.N. Islam, Ajit Gupta and Ranjit Kumar Paul

[1]International Maize and Wheat Improvement Center (CIMMYT)
[2]Indian Agricultural Statistics Research Institute (IASRI), New Delhi

E-mail: a.rai@cgiar.org; a.ajay@cgiar.org

Technologies advancements are bringing great changes in the area of data collection, storage and analysis. Digital data collection (DDC) is a process of collecting data electronically using smart phones, tablets and net-books. This version has significantly improved data quality and reduced resource requirement for field surveys in past few years. Several DDC tools (Kobo, Collect, SurveyCTO, Magpi, Cogo, Insyt, GoSurvey, etc.) have been recently developed and are available for use. Each of them is having its own benefits and limitations in the context of utility and deployment by users. Cereal System Initiative for South Asia (CSISA) in collaboration with Indian Council of Agriculture Research (ICAR) in India used Open Data Kit (ODK) for diagnostic survey of crop production practices. In 2018, the survey was implemented in 50 districts across five eastern states (Bihar, Uttar Pradesh, Odisha, West Bengal and Chhattisgarh) through Krishi Vigyan Kendra (KVK) of each district. The survey intended to capture detail information from rice-wheat cropping system so the focus was

more on Bihar and east part of Uttar Pradesh along eastern Indo-Gangetic plain. From each district, 210 randomly selected farmers were interviewed in depth for their current crop production practices; approximately 8,000 data points were gathered. The objectives of this electronically enabled survey were to fill existing data gaps, generate recent data-based evidences, derive better insights and facilitate informed decisions by policy makers.

About ODK

Developers and researchers at Department of Computer Science and Engineering, University of Washington had founded ODK. ODK began as a Google sponsored sabbatical project in April of 2008. The first two deployments of the tool happened in Uganda and Brazil (https://docs.opendatakit.org/). ODK is an open-source tool the source code is available for free and is licensed to permit customization by users. These are generally developed as a public collaboration and made freely available. Compared to conventional paper based data collection, ODK provides great ease by automating data compilation. In large scale survey, data compilation itself require huge resources and task is very much error-prone, whereas, ODK was easy to use and easy to scale even in resource-constrained environments.

There are three major components (Build, Collect & Aggregate) that jointly form the data ecosystem in ODK.

ODK Build: This is used for designing a questionnaire for ODK.
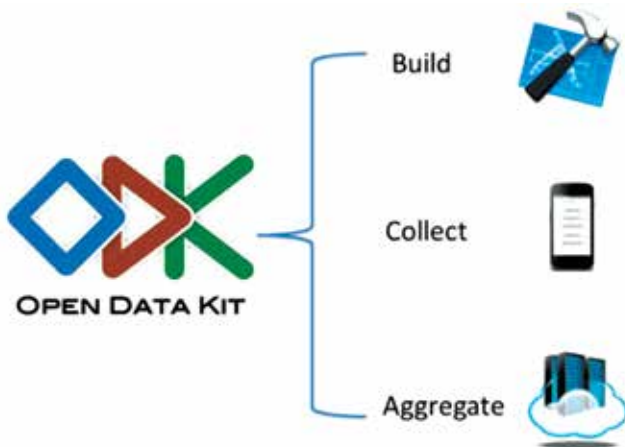
ODK Build is a form designer with a drag-and-drop user interface. Build is an HTML web application and works best for designing simple forms. Alternatively, XLSForm is a form standard created to help simplify the authoring of forms in Excel. XLSForms are simple to get started with but allow for the authoring of complex forms. Forms designed with Excel can be converted to XForms that can be used with ODK tools.

ODK Collect: It is an Android app that is used in survey-based data gathering. It supports a wide range of question and answer types, and is designed to work well without network connectivity. ODK Collect renders forms into a sequence of input prompts. Users work through the prompts and can save the submission at any point. Finalized submissions can be sent to a server. Collect supports location, audio, images, video, barcodes, signatures, multiple-choice, free text, and numeric answers.

ODK Aggregate: It is a Java application that stores, analyzes, and presents XForm survey data collected using ODK Collect. It supports a wide range of data types, and is designed to work well in any hosting environment. With Aggregate, data collection teams can:

- Host blank XForms used by ODK Collect
- Store and manage XForm submission data
- Visualize collected data using maps and simple graphs
- Export and publish data in a variety of formats

Accordingly, the workflow for data collection through ODK system is as follows:

- Design the form (questionnaire)
- Download a questionnaire for data collection
- Collect the data, **even if device is offline**
- Submit collected data to ODK Aggregate
- Access aggregated data for use

The mobile app i.e. ODK Collect, to be used by enumerators can be downloaded from Google Play Store. The updated version (v1.20.0) of the app contains six buttons and their functions are self-explanatory. Once the mobile app gets linked with the hosting server, these buttons rightly perform following functions.

- *Get Blank Form* – It is used to download desired survey forms in the data collection device from server. Internet connectivity is required.

- *Fill Blank Form* – It is used to fill-in the information in the form while conducting the survey. It works offline.

- *Edit Saved Form* – If enumerator wishes to add/change some information in the surveyed form before sending to the server, it can be saved in the device. This button can be used for doing edits.

- *Send Finalized Form* – It is used to send single or multiple surveyed forms from collection device to the server. Internet connectivity is required.

- *View Sent Form* – If you wish to see how many forms you have sent through a particular device, it generates the list of sent forms.

- *Delete Saved Form* – It is used to delete blank form, if the current form is obsolete or an updated version of blank form has to be used. This button can also be used to delete filled-in forms if users do not want to submit it on server. It mostly happens in case of form testing.
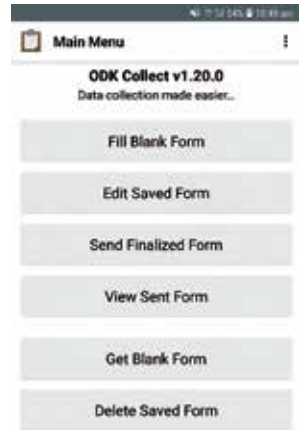
Benefits of ODK

There are several reasons for preferring ODK in the current landscape diagnostic survey. As the diagnostic survey is quite large in terms of sample size, spread and length of questionnaire, manual data compilation would have been extremely difficult to handle. The respondents of this survey are farmers and they are mostly located in hinterlands. So, we wanted a tool that can work uninterrupted in such setting. Another factor of choosing ODK was the confidence of CSISA's technical team in handling the tool. ODK had been used by CSISA for almost five years for collecting and monitoring data. Considering these factors, it was decided to go with ODK for the current landscape survey. In general, ODK provides other benefits over conventional paper based survey system. The key benefits are as follows:

Cost: There are many elements of cost. Electronic devices of course cost more than paper. But, when we factor in the requirement of hiring, training and employing data entry staff for the paper processes, in addition to buying and setting up the data entry machines, it ends up being costlier.

Speed and Efficiency: This is the most obvious advantage of digital data collection over paper-based system. Digital data collection reduces both data collection time and also the time required to analyze and distribute results. One of the main issues with paper version is its in-field administration if changes arise. While digital forms can be updated and pushed to enumerators quickly and automatically.

Data quality: Digital data collection reduces the possibility of error at the point of in-field collection, and it can also automate data correction. Paper can be lost, destroyed, or mishandled in a number of ways, which can create problems later if the data needs to be re-accessed. Digital data, on the other hand, can be easily and inexpensively stored, copied, and backed up.

Visibility and Tracking: Another important advantage of Digital Data Collection (DDC) is tracking. Paper process does not tell us anything about what is going on in real time, but with a digital platform, as soon as an enumerator completes and submits a form, the data is accessible to all stakeholders. We can check who has sent this, from where it has come and is there any discrepancy. Data managers can contact back the data collector in case of need.

Functionalities

ODK provides wide range of functionalities right at the time of questionnaire designing that improve data quality and restricts users to enter incorrect data. Some of these features are:

Skip patterns: Questions with skip patterns are very common in any form of survey. For example, we may only want to ask a respondent about irrigation frequency, if their response to a previous question on whether they have irrigation facility is "yes". These types of skip patterns can only be enforced on digital surveys, with a conditional question only appearing based on the response to a previous question. An example of a skip pattern question is as below:

Do you have irrigation facility?   [   ] Yes / [   ] No

If Yes to question above, how many times you irrigated your crop   [Number Entry_____]

For paper based questionnaires, proper recording of such skip pattern kinds of questions are entirely reliant on the enumerator skills, knowledge of the questionnaire and keenness, leaving plenty of room for error.

Entry limits: This kind of restriction is usually vital especially for numeric types of questions. For digital surveys, it is possible to restrict entries, by having minimum and maximum values. For example, when taking the second split of urea applied in days after seeding, it cannot be less than the value of days (10-30) entered for first split. We can restrict conditional entry to higher value of first split in days. Any entry below that is therefore rejected.

Type questions: Survey questions happen to be of different types. These can be numeric, alpha-numeric, and dates, among other types. ODK ensures that entries are limited to their type, so we don't have a text response for a numeric question. Form developer is also able to control date format through pop-up calendar, furnishing options as single select or multiple select, pre-populating basic information such as area details, etc.

Optional vs mandatory questions: In DDC, we have control over whether a question is mandatory or optional. In this case, enumerator does not miss responses for questions that are considered essential for the survey. For example, you cannot move forward with the interview unless you fill the response about variety type. This means that the data available for analysis is usually pretty clean and ready for analysis.

Geo-Tagging

One of the best features of ODK-based survey is geo-referencing. Currently available mobile hand-sets can capture geo-location even without having internet and mobile connectivity. It adds great credibility in data we collect through ODK. All the locations (largest plot of respondents) of landscape diagnostic survey henceforth are geo-tagged. It further allows us to layer this data with other parameters such as, soil profile, weather condition, etc. Geo-points of

respondents of Bihar and eastern Uttar Pradesh have been furnished below:
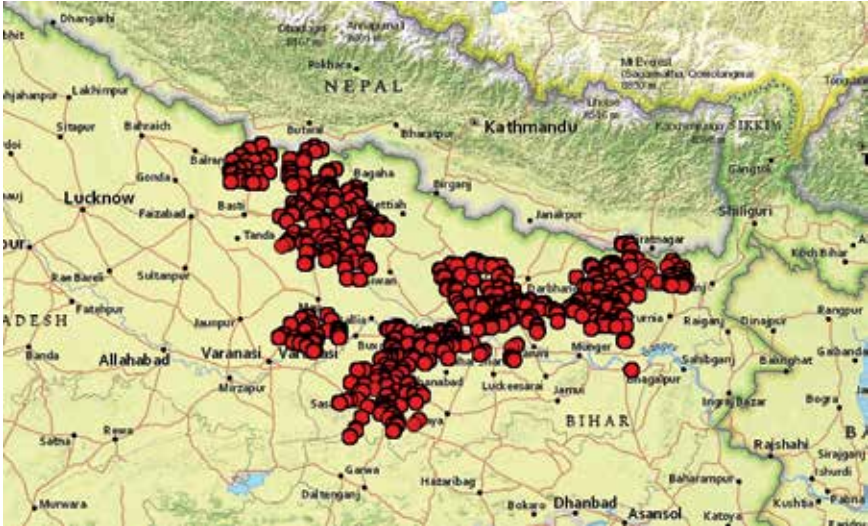
Reference

OPEN DATA KIT Documentation. (2017). Retrieved from ODK: https://docs.opendatakit.org/