New Frontiers in Agricultural Extension – Volume II

Editors

A.K. Singh, Peter Craufurd, Sudhanshu Singh, Virender Kumar, Balwinder Singh, Andrew McDonald, Ajoy Kumar Singh, Anjani Kumar, Randhir Singh, Rajbir Singh, Prakashan C. Veettil, P. Panneerselvam and R.K. Malik

The Cereal Systems Initiative for South Asia (CSISA) is a regional initiative to sustainably increase the productivity of cereal-based cropping systems, thus improving food security and farmers' livelihoods in Bangladesh, India and Nepal. CSISA works with public and private partners to support the widespread adoption of resource conserving and climate resilient farming technologies and practices. The initiative is led by the International Maize and Wheat Improvement Center (CIMMYT), is jointly implemented with the International Food Policy Research Institute (IFPRI), the International Rice Research Institute (IRRI) and Indian Council of Agricultural Research (ICAR) and is funded by USAID and the Bill & Melinda Gates Foundation.

www.csisa.org

Email: cimmyt-csisa@cgiar.org

***Citation*:** A.K. Singh, Peter Craufurd, Sudhanshu Singh, Virender Kumar, Balwinder Singh, Andrew McDonald, Ajoy Kumar Singh, Anjani Kumar, Randhir Singh, Rajbir Singh, Prakashan C. Veettil, P. Panneerselvam and R.K. Malik (Eds.) 2020. **New Frontiers in Agricultural Extension - Volume II**. International Maize and Wheat Improvement Center (CIMMYT). pp. 550.

This publication was made possible through the support provided by the United States Agency for International Development and the Bill & Melinda Gates Foundation. The contents and opinions expressed herein are those of the Editors and do not necessarily reflect the views of USAID or the Gates Foundation. The support from Indian Council of Agricultural Research is acknowledged.

Published in December 2020

New Frontiers in Agricultural Extension

– Volume II

Contents

1.2 Application of open data kit in landscape diagnostic survey for rice in India

A. Ajay[1], S. Sharma[1], A. Samaddar[2], A. Arora[3], S. Pal[3], S. Marwaha[3] and S.N. Islam[3]

[1]International Maize and Wheat Improvement Center (CIMMYT), Patna, Bihar
[2]International Rice Research Institute (IRRI), New Delhi
[3]ICAR-Indian Agricultural Statistics Research Institute (IASRI), New Delhi

E-mail: a.ajay@cgiar.org

Technology advancements are bringing a great change in 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.) were 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 Agricultural Research (ICAR)-Indian Agricultural Statistics Research Institute (IASRI) in India used Open Data Kit (ODK) for diagnostic survey of rice production practices. In 2019, the survey was conducted in 70 districts across nine Indian states (Bihar, Uttar Pradesh, Odisha, Andhra Pradesh, West Bengal, Chhattisgarh, Jharkhand, Haryana and Punjab) through *Krishi Vigyan Kendra* (KVK) of each district. The survey intended to capture detail information on current rice production practices being applied by farmers. From each district, 210

randomly selected farmers were interviewed in depth for their rice crop production practices; approximately 14,000 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 Open Data Kit (ODK)

Developers and researchers at Department of Computer Science and Engineering, University of Washington had developed ODK. ODK began as a Google sponsored sabbatical project in April 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 a large scale survey, data compilation itself requires huge resources, and the task is very much error-prone; whereas, the ODK is 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 (Fig. 1).

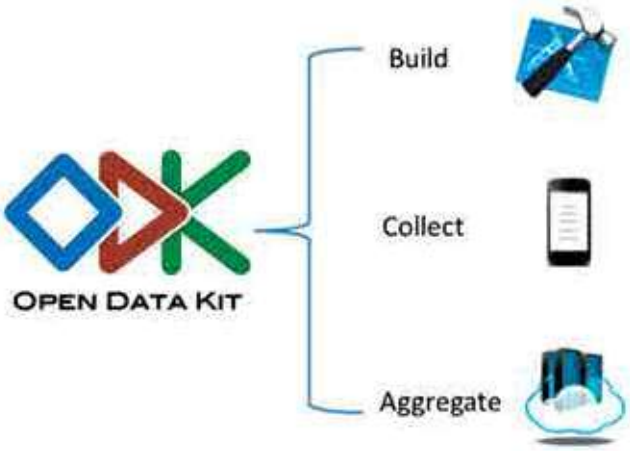


Fig. 1. Components of ODK platform.

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, the .xml version readable by the ODK server as digital questionnaire and used by ODK Collect.

ODK Collect

It is an Android app that is used in survey-based data gathering. It supports a wide range of question and answers 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 :

i. Design the form (questionnaire)

ii. Download a questionnaire for data collection

iii. Collect the data, **even if device is offline**

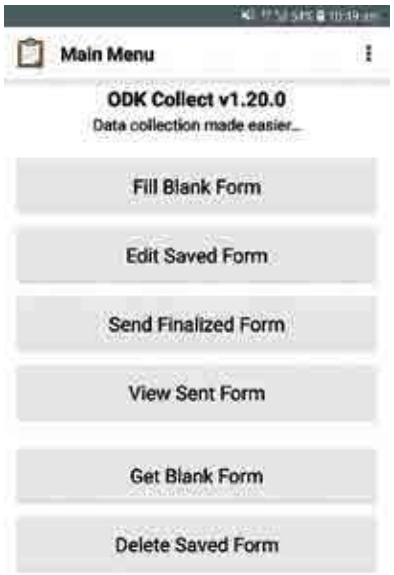
iv. Submit collected data to ODK Aggregate

v. 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 (v1.29.1) 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 were several reasons for preferring ODK in the current rice production practices 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. CSISA had been using ODK for almost five years for collecting monitoring data. Considering these factors, it was decided to go with ODK for the current landscape survey. In general, ODK provides another benefits over conventional paper based survey system. The key benefits are as follows :

Cost : There are many elements of cost. Electronic devices 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 for analyzing and distributing 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 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 need to be re-accessed. Digital data, on the other hand, can be easily and inexpensively stored, copied, backed up.

Visibility and tracking : Another important advantage of digital data collection 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 are 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 if required.

Functionalities

ODK provides wide range of functionalities right at the time of questionnaire designing that improve data quality and restrict users to enter incorrect data (ODK Documentation, 2017). Some of these features are discussed here.

Skip patterns : A question with skip patterns is very common in any form of survey. For example, we may only want to ask respondents 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 is 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 of 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 do not 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 digital data collection, 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 are 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 to data we collect through ODK. All the locations (largest plot of respondents) of production practice survey

Operational model of ODK has been furnished in following schematic diagram

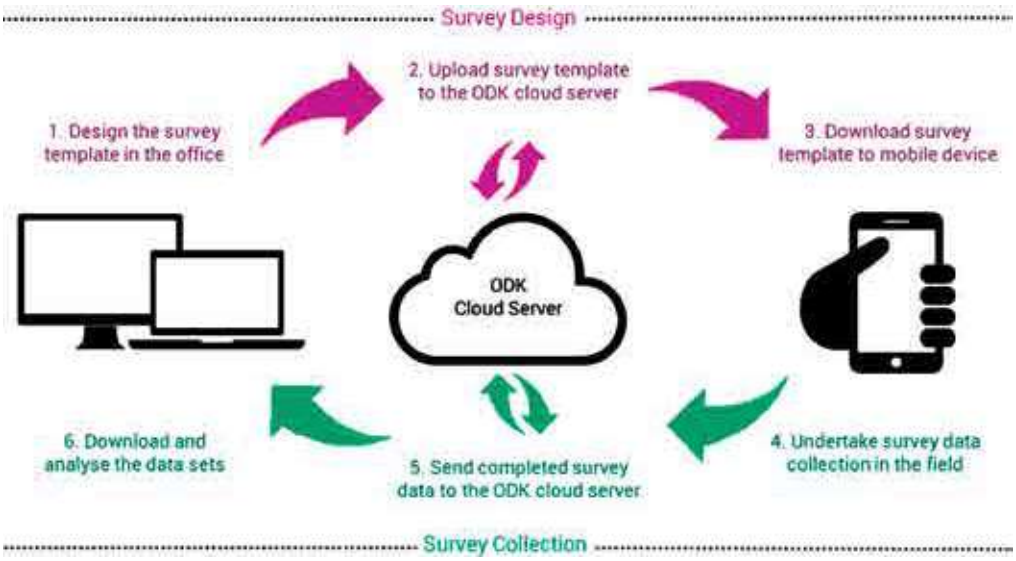


Fig. 2. Operational model of ODK.

henceforth are geo-tagged. It further allows us to layer this data with other parameters such as, soil profile, weather condition, etc.

Extracting of dataset : After data collection is complete, the compiled dataset of the survey is extracted from the ODK server. ODK Aggregate component is used for data transfer on the server. A server has been hosted in ICAR Data Center at ICAR-IASRI with the domain https://csd.icar.gov.in/, to transfer the curated data collected in surveys in the Indian region. Data extraction, a simple and easily executable task, can be performed any time after the collection process starts. Personnel having access of server should log-in and choose dataset from the FORM box option as shown on the left of Fig. 3. Once a particular dataset (Form) is selected, the user has to tap on EXPORT button as depicted on the right side of Fig. 3. ODK server starts generating desired dataset and the same can be downloaded. Data can be extracted in excel file and the same can be shared using

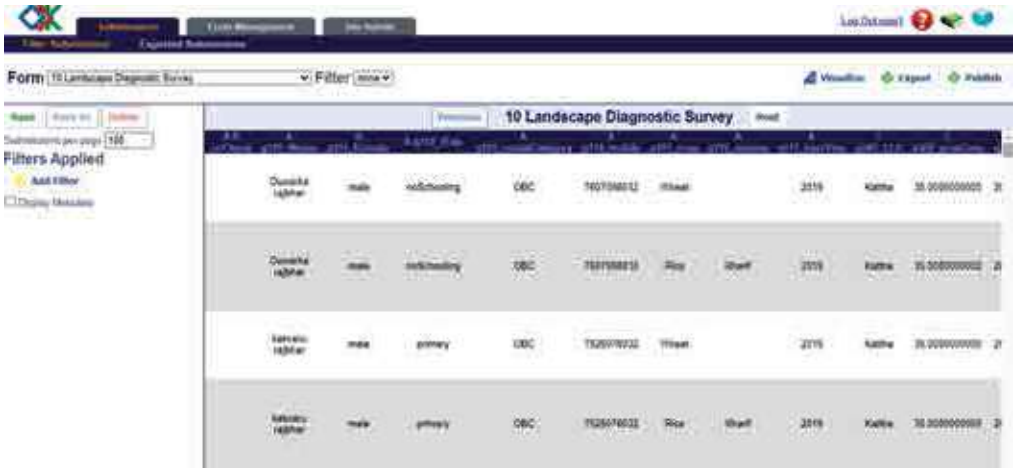


Fig. 3. Snapshot of ODK server showing aggregated data points of individual surveys in rows.

email, Google Drive or Dropbox. Data can be transferred to database for further processing.

Data diagnostics : ODK Aggregate provides pie chart, bar graph and map to diagnose the datasets. In LDS survey, user can see the distribution of dataset with respect to states, crop and other attributes. An example of diagnostics is shown for distribution of data from different states in Fig. 4. This type of data diagnosis helps in understanding the larger picture of the dataset. Such data driven findings can also form the basis of refining analytics and help to streamline the internal learning process.

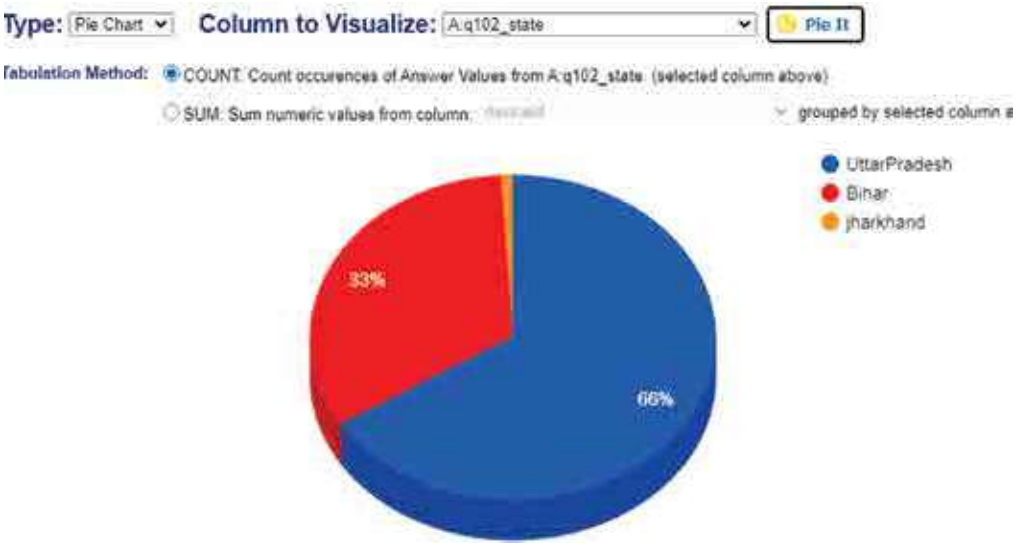


Fig. 4. Data visualization on ODK server.

Conclusion

ODK (Design, Collect and Aggregate) is completely self-reliant system for mobile based surveys. The surveys can be carried out in an efficient and affluence way using ODK. Data can be collected using mobile devices and easily converted to an excel file or seamlessly integrated with database. This makes it possible to deliver results of complex surveys in minimized time frame.

References

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

ICAR-IASRI (2019). ODK Server. Retrieved from https://csd.icar.gov.in