13

TRANSFORMING RURAL DEVELOPMENT THROUGH OPEN DATA KIT (ODK): APPLICATIONS AND IMPACT

H. K. Solanki

Assistant Professor (Sr.), Centre for Rural Infrastructure, National Institute of Rural Development and Panchayati Raj, Ministry of Rural Development, Government of India), Hyderabad, India
Email: hksolanki.nird@gov.in

Abstract: *The National Institute of Rural Development and Panchayati Raj (NIRDPR), Ministry of Rural Development, Government of India, is an apex institute of the country for training, research and consultancy in rural development. With the support of OSGeo, a global non-profit organisation supporting open-source geospatial software, and its India chapter, NIRDPR has initiated using the Open Data Kit (ODK) in research studies and projects. The objective of using ODK in NIRDPR is to save cost, time, and effort in research studies using a reliable, open-source, and easy-to-understand mobile-based data collection tool. Researchers at NIRDPR are using ODK in key thematic areas like rural housing, rural roads, skill, sanitation, etc, under various relevant flagship schemes of the Ministry. NIRDPR completed many studies with the help of ODK-based field data collection and aggregation. NIRDPR is an ideal organisation for propagating ODK and other open-source tools to the grassroots level for societal benefits. It is expected to adopt ODK 2.x tool suit for future improvements over some present constraints.*

Key words: NIRDPR, Open Source, Open Data Kit, ODK, Rural Development, OSGeo, Transifex

Introduction

Collecting, storing, and managing field data are essential requirements in field research and project execution in various sectors. In the rural development sector, especially in poor or developing countries, there is a need for technically sound and scalable data collection and management solutions at low cost or no cost. "Open Data Kit (or ODK for short) is an open-source suite of tools that helps organisations author, collect, and manage mobile data collection solutions" (ODK Community, 2018-a, para. 1). ODK is a community-driven tool that produces free and open-source software for collecting, managing, and using data in resource-constrained environments (ODK Community, 2018-b). The operation of ODK is divided into three parts: *Build*, *Collect* and *Aggregate*. The Build part comprises the generation of questionnaires and forms in ODK-compatible formats using various tools like websites, offline software, MS Excel, etc. The collect segment mainly comprises mobile operations, including data collection, storing and managing data in mobiles and sending to the server or extracting data offline directly from mobile. The aggregate segment aggregates the data in the online server, which may be operated on laptop or desktop computers, depending on available resources and the extent of use.

ODK is helpful in other related sectors like agriculture, community health, disaster response, and environmental conservation. However, the article will describe its practical applications in the rural development sector at the National Institute of Rural Development and Panchayati Raj (NIRDPR), Hyderabad, India. NIRDPR is an apex organisation of the Ministry of Rural Development, Government of India, to cater to the training and research needs of the rural development sector. NIRDPR is also involved in consultancy projects for the government of India, state governments, and other organisations. NIRDPR works in close coordination with the Indian Technical and Economic Cooperation Programme (ITEC) (Ministry of External Affairs, 2015), African-Asian Rural Development Organization (AARDO, n.d.), Centre on Integrated Rural Development for Asia and the Pacific (CIRDAP, 2018) countries for fulfilling their training and research needs in Rural Development sector. NIRDPR proliferates these countries' latest tools, techniques and best practices.

Use of ODK in Rural Development

ODK users include Google, the World Health Organisation (WHO), the Centre for Disease Control and Prevention (CDC), the United States Agency for International Development (USAID), the Red Cross and Red Crescent, the Carter Center (The Carter Centre, 2016), the Jane Goodall Institute, cited at http://opendatakit.org/. Other studies are described and showcased in *the showcase* section on the ODK forum website at https://forum.opendatakit.org/c/showcase. A detailed list of other research works in the form of papers, talks, posters, and demos is available in *the research* section of the official website of ODK at https://opendatakit.org/community/research/. A detailed list of press releases and country-specific published use cases of ODK is available in *the history* section of the ODK website at https://opendatakit.org/community/history/.

In India, organisations using ODK in their research studies or projects independently or in collaboration are described further, highlighting the collaborative nature of ODK's use and making the audience feel connected and part of a larger community. Under the Performance Monitoring and Accountability 2020 (PMA2020) project of the Bill & Melinda Gates Institute for Population and Reproductive Health, Department of Population, Family and Reproductive Health, Johns Hopkins Bloomberg School of Public Health, ODK is being used extensively to collect data on various indicators related to hygiene, sanitation etc from countries of Asia and

Africa. The project is operational in 11 countries through a network of universities and research institutes (Johns Hopkins University, 2018). Indian Institute of Health Management Research (IIHMR) is India's implementation partner for this project (IIHMR University, 2018). ODK has been used in the Local Self Government Department, Govt. of Kerala, India, in District Thrissur on a pilot basis for mapping the Local Body Institutions. It is found suitable to be replicated in all Government Departments of Kerala State in future (ODK Forum, 2017-a) Regional Remote Sensing Service Centre (RRSSC). Indian Space Research Organisation (ISRO), Jodhpur, is using ODK as a project technology partner for survey work in a project of the Ministry of Health and Family Welfare, Government of India, showcasing the diverse applications of ODK and inspiring the audience with its versatility.

The Foundation for Ecological Security (FES), a leading non-governmental organisation, uses the tool extensively in its projects. NIRDPR has also taken advantage of their experience using the ODK tool. Individuals of Indian origin based in India or outside, namely Rohit Chaudhri (ODK Community, 2018-c, para. 8), Shobhit Agarwal (ODK Community, 2018-d, para. 6), Akshay Patel, and Narendra Singh (ODK Forum, 2017-b) are working in core development groups and actively engaged with official ODK Forum and contributing to the development of the ODK tool in various capacities. NIRDPR has conducted two practical trainings of five days duration in the financial year 2018-19 for Forest officers of the Government of Rajasthan on 'Open-Source GIS Tools for Forest Resources Management' in collaboration with the Department of Forests, Government of Rajasthan. In both the training, ODK introduction along with hands-on field data collection and data management were part of the training contents. The Forest Department of the Government of Rajasthan uses ODK in various forest surveys of different scales. The uses of ODK described above are selected and limited to the interactions and knowledge of an author or showcase section of the ODK forum in the Indian context for the rural development sector. Other tools of the ODK ecosystem, like *KoBoToolbox*, are also being used by various organisations and individuals relevant to the Rural Development sector.

Operationalisation of ODK in NIRDPR

Data related to research studies and consultancy projects in NIRDPR were collected in paper-based survey forms. It led to more time, energy consumption, and possible errors during manual data entries. It was initially thought at higher administration to have a dedicated in-house solution with its resources of NIRDPR. With the author's initiation, it was decided to use ODK in NIRDPR so that existing widely used open-source tools and technologies can be used with the possibility of future need-based customisation. To train the NIRDPR faculty and research staff on the tool and for the installation of the local ODK server, a three-day workshop was planned with the help of the OSGeo team. OSGeo India chapter was keen and ready to provide technical support in the training workshop and installation of server computers.

Three days ODK Workshops in NIRDPR

A three-day Open Data Kit (ODK) workshop was organised from 2-4 November 2017 in NIRDPR, Hyderabad. The first two days of the workshop mainly concentrated on introducing and using ODK tools like building a form, uploading to the server, downloading to mobile, collecting field data, sending the collected data to the server, and downloading it for analysis. Initially, participants created the forms individually, then were divided into groups, and they prepared the refined forms per their work areas. The collected data were also demonstrated and plotted using Open Source QGIS software for spatial visualisation. The method to prepare multi-lingual forms and create forms using Excel sheets in offline mode was also described. The participants could create the forms and upload/download the forms from the server and

data collection. The participants collected sample data from the NIRDPR campus. The workshop was unique to the institute in providing an overview of the technology. It made the tools and techniques available for data collection, visualisation, and analysis, as well as the setting up of a dedicated server. On the third day of the workshop, the ODK local server was configured in NIRDPR on a dedicated computer by OSGeo India experts. The server was tested well by IT staff and faculty members of NIRDPR on the same day. Resource persons Dr. Venkatesh Raghvan, Professor, Osaka City University Japan; Shri. Ravi Vundavalli, Ex-Director, Geological Survey of India & Secretary OSGeo India; Shri Natraj Vaddadi, Charter member, OSGeo Foundation; Shri. Ramamurthy, Ex-Director, Geological Survey of India & Charter member, OSGeo Foundation; Shri. V. Balasankar, Faculty member, Aditya College of Engineering, Andhra Pradesh, author, delivered the lectures, hands-on sessions, field data collections, server configuration and testing, etc. The workshop was attended by 61 participants from NIRDPR, including faculty members and research staff of various centres of NIRDPR. After the workshop, most of the field data collection under research studies and consultancy projects of NIRDPR are done through ODK.

Workflow on ODK in NIRDPR

PostgreSQL is used for the database on the server computer, and Ubuntu 16.04 is used as an operating system. A core team at the Centre for Information and Communication Technology (CICT), NIRDPR, is working towards managing ODK data and workflow in NIRDPR.

User Rights in ODK

ODK, in its original form, gives four kinds of rights to users: 1) Data Collector, using these rights and related credentials, can only download the survey forms and send the data to the ODK server. Further, data collectors can locally manage the data on mobile by editing, deleting, etc. However, data collectors can extract their data from mobile devices and copy it to a desktop or other media. From there, it can be further extracted in usable formats (.csv and media) using *ODK Briefcase* utility (Open Data Kit, 2017); 2) Data Viewer- using this proper and related credentials user can view the submissions and export the data to usable formats (.csv, .kml and JSON) without the possibility of deleting any data or form at Aggregate site; 3) Form Manager- using this right, in addition to the rights of Data Viewer user can upload new ODK compatible forms in .XML format and can delete the submitted individual response or complete forms; 4) Site Administrator- this right is highest right in ODK and in addition to the rights of Form Manager, user can add new users/credentials with restricting the rights of them up to any level. The site Administrator can also delete the previous users and credentials. As per research ethics and for data security, research teams of NIRDPR, including faculty members, are provided with Data Collector rights only. The core team of CICT is managing other rights.

Data and Process Management

Based on the guidance of the core CICT team, research team members convert a research study's survey schedules/forms to ODK-compatible forms in XML formats. This is done on the ODK Build site (http://build.opendatakit.org) with individual credentials. For complex forms, *XLS Form* is used. "XLS Form is a form standard created to help simplify the authoring of forms in Excel" (http://xlsform.org/en/). Each Excel workbook usually has two worksheets: *survey* and *choices*. A third optional worksheet called *settings* can add additional specifications to the form. Additionally, the http://opendatakit.org/xlsform/ site provides a platform for online testing of these Excel sheets for any syntax error. An offline tool can also be downloaded from this site to convert Excel to XML format. On this site, one more open-source utility, 'Enketo,' may be used to preview the created form on the computer.

Once the form is generated and converted to .XML using the above online or offline tools, these .XML files are copied from the PC and pasted into the *Forms* folder inside the ODK folder available in the internal memory of mobile if ODK Collect is installed already. Then, this form is visible in *the ODK Collect* mobile app. Then, ODK forms are tested on mobiles, taking dummy data or doing a pilot survey to test schedules or questionnaires. If any correction is necessary after testing, the original blank form and data collected are deleted using the 'Delete Saved Form' option in *the ODK Collect* app. After sufficient iterations, the final XML form is sent to the CICT team through email for uploading to the Aggregate server for further download in mobiles for actual use. From the ODK server, data collectors or field investigators download the forms of relevant studies using the *Data Collector* credentials and *the Get Blank Form* function in the app. After data collection, they can send data from mobiles to Aggregate after the survey. During the survey, internet connectivity or a mobile network is not required. Mobiles can also be put on aeroplane mode to save battery life. Internet connectivity is required to send data to the server computer.

The research coordinator consolidates the data from research studies or projects from time to time. This data is supplied by the CICT team using *ODK Briefcase* utility in the form of a *.csv* file containing responses and *a media* folder which contains photographs/media. Only image files are suggested for each response, considering the limitation of internet speed. Questions related to taking video and audio files are avoided. Once the survey and fieldwork are declared as completed, final data in the form of .csv and media is provided to the research coordinator or principal investigator. This data is used in a GIS environment using QGIS. Other statistical analyses are done using different statistical software based on the choice and availability of software with the research team. Fig 3 below depicts this workflow.

Figure 01: ODK Operationalisation and Workflow in NIRDPR
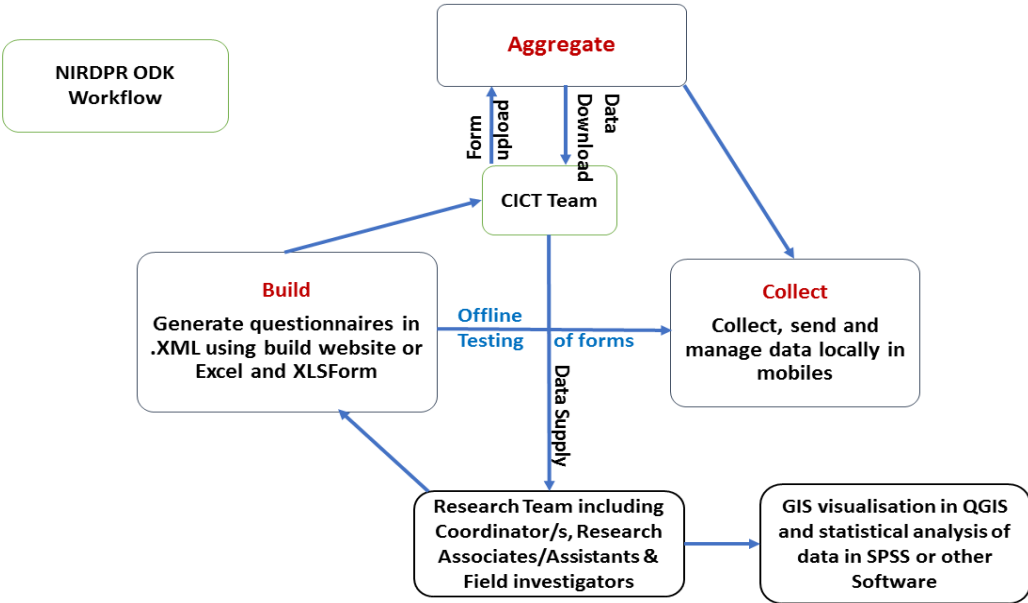
Achievements and Impact

In NIRDPR research studies use of ODK has been mandated. Field data collection in a few Centres for Geoinformatics Applications in Rural Development (CGARD) consultancy projects is also done on ODK. Until now, the studies and projects using ODK-based data collection include thematic areas/schemes of Rural Roads, Pradhan Mantri Awas Yojna (PMAY), Sanitation, Open Defecation, Manual Scavenging, and Skills. The data collection part under these studies was completed much faster with nil or fewer manual mistakes, which ordinarily happens during data transfer, conversion and entering into systems from paper-based surveys.

More studies are coming on the ODK system. ODK has become an essential part of all field-based research studies on NIRDPR. Besides using the tool in research studies and projects, CGARD and NIRDPR have started introducing it in their regular national and international training programmes.

Language Localisation of ODK in Hindi
Open Source gives users the freedom to be involved in all spheres of the development cycle of a project. Getting involved in language localisation is also an opportunity for the users. Taking the experience of the author's involvement in the language localisation of QGIS software in Hindi, the ODK interface has also been translated into Hindi at NIRDPR using the official online Transifex (Transifex, 2018) platform. For translation of the ODK interface, survey forms and entering the survey responses in different languages, 'Google Translate' and 'Google Input Tools utilities are being used.

Findings
The tool has been widely accepted by NIRDPR faculty and research staff and is being used in research studies and projects. With improvement in knowledge and practice, more confidence is raised in the general user community of the organisation. The tool has been considered helpful compared to commercial tools, seeing the possibility of customisation and in-house availability of knowledge resources. Based on the internal discussions with users of the tool within the organisation, it has been observed that mobile-based data collection tools are more suitable for data collection when the number of variables is less and the number of responses and/or geographical extent is more significant. More swipes are to be done in more variables, and fewer respondents limit the tool's usability commensurate to the efforts made on developing and designing ODK-compatible forms compared to paper forms. What are the optimal number of questions/variables suitable for mobile surveys, and how many respondents are these kinds of tools suitable? Require further research in this direction. However, when research ethics and data validation are a concern, the application of ODK is always advantageous. It is observed that the research team requires more command over workflow. Due to ODK 1.x limitations, data viewer and above rights are not provided to research teams; they can only send and collect the data as depicted in Figure 3 above in the workflow. The use of ODK 2.x tool suit in future above limitations may be minimised. For other limitations, engaging dedicated and appropriate staff with programming knowledge may help customise the tool per local needs. For the development of complex forms, engaging an additional expert or the same expert with proper knowledge of Excel and data management for exclusive support to research staff in organisations like NIRDPR is necessary.

Future Directions
With the increase in the number of studies, a dedicated ODK administrator/team may be hired to give expert guidance on making complex survey forms and managing research data and data analytics. The team may scale up or customise the operations of ODK for a seamless workflow flow to reduce the tasks of research teams. This may include giving the researchers the facility to provide form-specific credentials to have complete control over the form and data without interfering with or viewing the data of other studies. Based on data flow and quantum of data, the hardware may be scaled up to have more storage capacities. Based on the quantum of work, purchasing tablets for larger screen sizes may also increase the interest and speed of work. In NIRDPR, ODK 1.x is presently being used, including Collect, Aggregate, XLS Form, Build, and Briefcase utilities. Most future requirements can be solved by adopting the new ODK 2.x tool suit, a recently released toolset for advanced uses when organisations want to scale up for higher operations (ODK Community, 2018-f) (Open Data Kit, 2017-b). As

NIRDPR has officially started using ODK for its research studies, migration to higher tools and exploiting the tool's full potential is important. With dedicated and appropriate manpower, installation, and use of ODK, 2.x tool will be advantageous for the future research environment of NIRDPR and rural development. A data visualisation platform linked with ODK Aggregate is also conceptualised for the future to see data on various online base maps for a broader understanding of data. Capacity building is necessary to proliferate the tool in the rural development communities and related organisations, and specific and dedicated training programmes of short or long durations from users' or developers' perspectives are required for the application of the tool in various thematic areas of rural development for the overall benefit of the community. NIRDPR is dedicated to achieving the above future possibilities, and the author, with the support of NIRDPR administration and staff, is pursuing the objective of broader application of tools for societal benefits in NIRDPR and in the rural development sector in general.

Conclusion

In developing countries' resource-constrained environments, data collection and management are crucial. In that scenario, having easy, readily customisable, scalable, free and open-source tools is a boon for rural development and developing countries. Organisations like NIRDPR, being pivotal and in the lead role in the development sector, can be ideal for proliferating the technology and knowledge to the relevant organisations of India and developing countries. Apart from training programmes conducted for countries associated with AARDO, CIRDAP and ITEC, NIRDPR is closely associated at national level with 29 State level institutes called State Institute of Rural Development (SIRDs) and 90 Regional Institutes under SIRRDs called Extension Training Centres (ETCs). ODK can be adopted by these institutions based on NIRDPR experiences and with the help of NIRDPR capacity-building programmes. NIRDPR is widely promoting the tool. NIRDPR supports and promotes open source and actively uses other open-source GIS tools, like QGIS, in its regular training and research.

References
1. AARDO, n.d., Organisational Structure-AARDO Members, Retrieved from http://www.aardo.org/aardohomepage/English/organizational-structure.html
2. CIRDAP, 2018, Member Country of CIRDAP-IINDIA-Link Institutions, Retrieved from https://cirdap.org/india/
3. IIHMR University, 2018, Performance Monitoring and Accountability 2020, Retrieved from https://www.iihmr.edu.in/stories/pma-2020
4. Johns Hopkins University, 2018, About PMA2020, Retrieved from https://www.pma2020.org/about-pma2020
5. Ministry of External Affairs, 2015, National Institute Of Rural Development – ITEC, Retrieved from https://www.itecgoi.in/courses_list.php?salt3=c6c1d681a02018-2019&salt4=f242edbf0935&salt=2dad16887b87
6. ODK Community, 2018, Open Data Kit History. Retrieved from https://opendatakit.org/community/history/
7. Open Data Kit, 2017a, Open Data Kit Documentation- ODK Briefcase, Retrieved from https://docs.opendatakit.org/briefcase-intro/
8. Open Data Kit, 2017b, Open Data Kit Documentation- Welcome to Open Data Kit 2's documentation, Retrieved from https://docs.opendatakit.org/briefcase-intro/
9. ODK Forum, 2017a, Open Data Kit in Local Self Government Department, Govt. of Kerala, India, Retrieved from https://forum.opendatakit.org/t/open-data-kit-in-local-self-government-department-govt-of-kerala-india-to/10517
10. Transifex, 2018, Dashboard-ODK Collect. Retrieved from https://www.transifex.com/opendatakit/collect/dashboard/