ORIGINAL RESEARCH PAPER

VOLUME-6 | ISSUE-5 | MAY - 2017 | ISSN No 2277 - 8179 | IF : 4.176 | IC Value : 78.46

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

A STUDY ON PAPERLESS SURVEYS: A GENERIC MOBILE APP FOR RESEARCH PROJECTS OF RURAL DEVELOPMENT DATA COLLECTION

Computer Science

Rama Krishna Thotapally Senior Programmer, Centre on Geo-Informatics Application in Rural Development, NIRD&PR

ABSTRACT

Information and communication technology (ICT) tools—hardware such as mobile phones, tablets and software applications such as Skype—are all around us: from the simple programme in our alarm clock to tablets and personal computers. Today, even orchards are managed by programmes that predict yields based on rainfall. ICT tools have revolutionised the way we work, especially how we deal with the mountains of data that accompany large-scale development projects. Today, there are more than 900 million mobile phones in India and even the poorest households have mobile access. Data collection for monitoring and evaluation (M&E) takes a great deal of time, staff hours and resources. When seeking to measure and compare the impact of interventions, quantitative data are very useful for providing hard evidence and demonstrating trends. Surveys are the most often chosen tool to collect these data. The idea was to try to simplify this massive undertaking by administering the surveys digitally, which allows for collection and processing of data in real time, while the survey was being conducted. The developed software application eliminated some of the conventional limitations associated with remote data collection and M&E but also introduced new challenges (e.g., no physical paper proof of the survey).

KEYWORDS:

Data Collection, Monitoring and Evaluation (M&E), Mobile Data Collection

INTRODUCTION

Managing survey data can be plagued by serious challenges at every step of data collection in the field, processing and analysis. Due to problems with storage and unauthorised access, data can be corrupted or completely lost. Furthermore, the sensitivity of personal information and privacy concerns make it a priority to ensure authorised access and prevent misuse.

With the help of Generic Mobile App for Rural Data Collection will support Authorised access – Only authorized users can access the compiled data and modify content through the webpage. Integrity and consistency – Every survey has a unique code. Upon completing a survey, a new code is automatically generated. The survey supervisor can monitor in real time each submitted form and the location of the surveyor. Data storage – As soon as the survey is uploaded, it is saved in a remote server. Automated analysis – Standardised graphs and charts for data analysis can be generated automatically with the software application.

The first step when the Researcher can create a research project information (project code, title, place and surveyor information) then surveyor visits the field is to enter the specified project and create fixed background information (country, state, district, category and year). Based on this information, the software application automatically assigns a unique ID number or code for each household. The starting time and household ID are locked and saved. The surveyor starts filling out the research project survey data by opening the form for the first indicator (e.g., land tenure). When the surveyor completes the first form, the second form is automatically opened and so on, until the last indicator form is completed. The surveyor can review the entire survey and send it to the server. The data for each form are saved by indicators in a separate line in the data table.

Each surveyor is required to log in before he or she can administer the survey, making it easy to trace and monitor progress. The fields for all the indicators are filled in by selecting the answer from a drop-down menu, which reduces data-typing errors. The application also captures additional information, including an image of the person or household being surveyed, the social and economic category, education level, number of family members, etc. Graphs and charts are generated after the household survey is completed and can be accessed through the web by authorised users.

The appropriate indicator for which the graph/chart is to be generated is selected, and the software application produces the outputs. Incomplete data sets are not included in the analysis but can be downloaded as raw data. There is less than 5% probability that the surveyor will not have mobile network coverage at the survey site. Should this problem arise, the survey can be completed in offline mode (using a form saved in TXT/XML format) and uploaded automatically when mobile network coverage is available. The process of conducting the survey can be monitored by the administrator in the project office (or remotely), and the information can be reviewed online. This allows real-time tracking[1] of results as they become available.

The administrator has the benefit of being able to check multiple operational and analytical aspects of the administration of the survey: who among the surveyors are currently online, which households are being interviewed, what are emerging trends in the survey and others.

DEVELOPMENT AND IMPLEMENTATION CHALLENGES:

Defining the unique code for each research project and surveyor - The administrator defines the log-in profile for each surveyor (name and email). When the surveyor conducts the survey, his or her name and email with household key is locked in a separate table. This prevents duplication as all keys are generated automatically.

f

Form compatibility across different mobile phone devices (i.e., standardisation of font, size, menus, etc.) -In the early versions of the software application, standard small style sheets were used, which were not compatible with different phone screen sizes. This problem was solved by employing HTML5 and CSS3 sheets for designing cross-compatible style sheets. Also, the design of the graphs and tables was fixed.

Language and handset compatibility -The free text inputs cannot be saved on the My-Sql tables, and several mobile devices and tablets could not support these fonts. To address this challenge, the fonts were converted to UTF8.

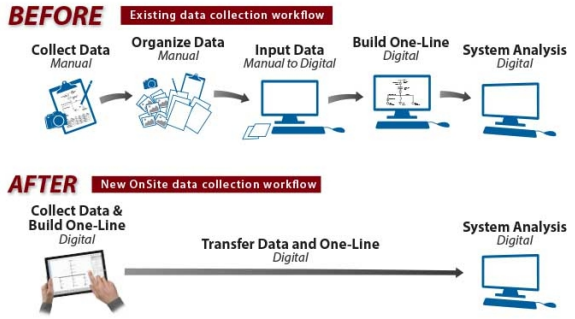

Fig 1: Differences of Manual Digital Data Collection

KEY BENEFITS:

- Time saved
- No typing errors in data entry
- Automated statistical analysis and generated graphs and charts
- Eco-friendly surveys: less transport, paper and electricity
- Real-time monitoring and evaluation
- Mobile phones can be used for other project related tasks

ROLE OF SMART MOBILES:

Data input or captured into phones may be transmitted or shared in many ways (including SMS, MMS, USSD, Bluetooth, wireless Internet, or the exchange of physical memory cards). Where mobile connectivity is not available, data can be stored on the phone and transmitted later once a phone is within sufficient range of a cell tower.

How and why might mobile phones
be useful in large-scale data collection efforts,
and what comparative advantages might their use have when compared to other options?

A number of attributes and characteristics of mobile phone use in such activities (as well as the use of other small, low-cost portable devices such as tablets, especially where such devices can be connected to mobile and wireless networks) may lead them to be considered, especially when compared with the use of more traditional, paper-based survey instruments:

Speed: Collecting data through the use of a mobile phone can greatly speed up the data collection process. Where network availability allows for near-instantaneous transmission of data to a central coordinating group, the reduced amount of time that elapses between local data collection and delivery can save weeks or even months of time in the overall data collection process. In addition, an early warning system of sorts can be established, allowing survey coordinators to quickly identify potential problems with data collection efforts and (potentially) correct them in almost real time.

Accuracy: Digital data capture at the source can greatly reduce transcription errors, and data transmission over mobile networks may ensure that no data are lost 'in transit'. Capture and transmission of data digitally may also ensure that it is easier to store and access them at later dates, should this be required.

Ubiquity, familiarity and convenience: Enumerators and survey respondents may, at a general level, already be quite comfortable using a mobile phone (and indeed may be using their own personal device), even if they have not used it specifically as part of data collection efforts. Such devices may be widely available already in target populations, who may be accustomed to their use in a variety of contexts.

Training: Because people may already know how to use the devices for many purposes, less technical training may be necessary in some circumstances. In addition, on smart phones, and to a lesser extent with feature phones, help files and on-screen prompts may provide useful relevant supporting documentation[2] and guidance that may reinforce messages from training that does occur, and potentially obviate the need for some sorts of training altogether.

Low power: Compared with devices such as laptops, mobile phones may be much easier to keep charged, as they require much less power and because many fast, low-cost charging options may be available in local communities because people are already utilizing such devices extensively for other purposes as part of their daily lives.

Combining with other data: Depending on the functionality of the phone used, textual data captured via mobile phone can be combined with data in other formats such as photographic images, audio, and video, as a way to substantiate the information provided by text. If a building being surveyed is noted as 'damaged', for example, an accompanying picture can provide further documentation. In addition, GPS or geo-locationary data can be passively collected and transmitted along with survey data. This can be used to help map the location of hospitals or schools, or to offer 'proof' that an enumerator actually visited a place that she has claimed to visit.

Low cost: All of these characteristics and affordances may mean that data collection enabled through the use of mobile phones can be done at substantially lower costs than is possible via traditional means.

It is worth noting that, in some circumstances, the comparison between data collection via traditional means and enabled through the use of mobile phones may be a false one. Indeed, in addition to improving the efficiency of data collection efforts when compared with traditional, largely paper-based practices, mobile data collection may also offer options for data collection that simply aren't feasible, or even possible, using other tools or methods.

ISSUES AND CHALLENGES:

While data collection via mobile phones may offer particular advantages when compared with traditional survey and data collection efforts, this is not to imply that such efforts are not without challenges or difficulties. Some common notable issues and challenges include those related to:

Technology: What technology should we use? What are the minimally viable specifications required for the devices used in mobile data collection efforts? What software applications are available, and what are their related advantages and disadvantages? These are often the first questions that many groups contemplating extending or complementing existing data collection efforts through the use of mobile phones often ask. While questions such as these will (or should) inevitably be posed at some point during the planning process, leading with them is typically a mistake. It is important that choices around a specific technology or device not define the initial scope or conception of the extent to which mobile data collection efforts might be beneficial and possible. No one technology may do everything that is required. Where one tool may appear to be a perfect fit for what is required, it may not work at the necessary scale. Vendors or partners may propose use of certain technologies or devices which may not be well suited for the specific data needs and/or data collection and sharing processes of the sponsoring group. Deciding which tools to use, and which partnerships to develop, may be critical pieces of the puzzle. First, however, it important to have a larger picture in place of what the objectives of a particular data collection might be, as well the relevant local context (including key stakeholder groups). Once that has been established, technology-related decisions can be made in the service of both higher order goals and on-the-ground realities.

Training: While in some instances less training and explicit instruction may be required, because the tool being used (a mobile phone) is familiar to respondents and enumerators alike, in some circumstances additional technology-related training and support

may still be required. The adoption of user-centered design principles and techniques can help in some circumstances to avoid the need for certain types of training. (If, for example, an interface is easy or even 'intuitive' to understand, users may not need to spend much time figuring it out.) That said, depending on the nature of the survey process and the methods of data transmission, additional training may well be required.

Cost: The costs of designing survey instruments delivered digitally may be considerably higher when constructing traditional paper-based questionnaires. In addition, new back-end technology infrastructure may need to be procured and put into place. Where it is not possible for enumerators to utilize their personal phones (whether because such use is not customary or permitted, because existing phones have insufficient functionality or because mechanisms for covering or reimbursing related data transmission costs are not in place), devices may need to be purchased and distributed (and potentially collected once they have served their purpose). Whether or not these costs in aggregate are cheaper than the way things are traditionally done or not – and they often are! – sponsoring groups may have difficulty estimating and budgeting for such costs on the front end, especially the first time they engage in efforts of this sort.

Data security: Digital collection and transmission of data as part of large scale survey efforts carries with it numerous potential risks and challenges related to data security and privacy that are often fundamentally different than those posed as a result of traditional paper-based survey efforts. If stored on devices, data can potentially be stolen or improperly accessed – the same holds true during data transmission. The use of encryption at both the device level and during transmission can greatly mitigate such risks, but the use of digital data security tools, protocols, and good practices, as well as related regulatory frameworks, laws and guidelines governing the use of such tools, may not be known widely known – or in some places, even be at hand. Where third party vendors or tools are utilized during the mobile data collection process – something that is usually the case – care needs to be taken to ensure that ownership, possession and utilization of data collected and transmitted are clearly articulated. Sufficient mechanisms need to be in place to audit related arrangements and agreements, and to ensure that penalties for non-compliance are clear and enforceable.

Speed of change: If mobile data collection efforts are expected to be repeated over time (for example, as a part of yearly census or annual planning activities), it should be expected that some of the enabling technologies may themselves change – as might the good practice guidelines and regulations related to their use. In some circumstances, this sort of change may even occur during the course of a specific data collection process itself! Technological change often outpaces the ability of planners and policy makers to anticipate and respond to these changes. Those involved in planning for and implementing mobile data collection activities would do well to keep this in mind, and should be prepared to monitor and respond to such changes over time. Planners should try to avoid making a 'big bet' on an unproven technology (especially one based on a closed or proprietary standard) or on single vendor, and should always consider how they may most effectively exit and transition from specific relationships, tools and standards.

MOBILE DATA COLLECTING SYSTEM:

The Mobile Data Collection [4] allows remote geographic locations collection and the transmission of data to central locations - data storing repositories through a wireless network. It is a combination of a client application running on mobile devices, wireless network infrastructure and remotely accessible database servers. Most of these systems share the same principles and guidelines to collect data remotely. As shown in Figure2, the process begins by designing a form, which contains a set of questions to collect relevant data. Collectors use these forms on the mobile device to collect real data in the field. The completed forms are stored on the mobile device until it's possible to upload them to the central server.

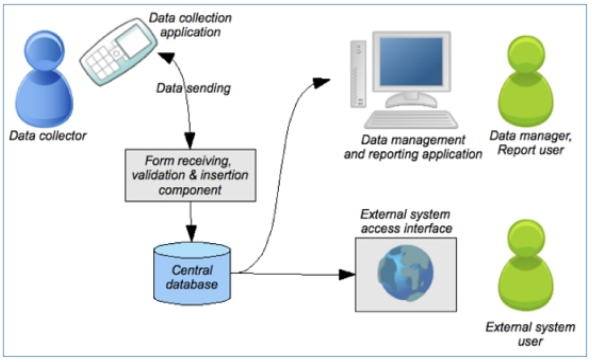

Fig2: Flow Diagram of Mobile Data Collection Process

PROPOSED ARCHITECTURE:

The goal of architecture for data collection in field research that works on heterogeneous computational environments is to disassociate the specificities of each collection with the used technologies, providing an architecture that can be used in various contexts for data collection. Subsequently, in this study, for a data collection architecture to work in several different computational environments this architecture must take two aspects into account: interoperability and flexibility. Interoperability is required so that architecture is not dependent on specific technology, but covering the possibility of using several technologies. Flexibility is required in two directions: the first to meet the diverse dimensions that data collection in the field can have, i.e., fulfill the data collection involving one or more institutions. A specific data collection can store the collected data in a single database located in a specific place or multiple databases located in several different places. The second is to make the necessary changes needed to pass the architecture to suit the contexts of data collection different, i.e., that the architecture is susceptible to changes - it can be configurable and reconfigurable. Architecture for data collection in field research that supports vector geographic data must use a DBMS with support for spatial data in both collection devices as the servers involved in collecting. Abstractly, the proposed architecture consists of three stages: the collection field, synchronization and storage, as shown in Figure 3. Each of these stages is described as follows.

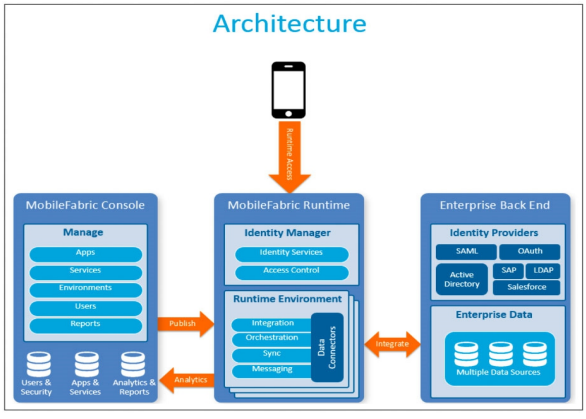

Fig3: Collection fields, synchronization and storage

The collection in the field occurs through a computer system in the mobile devices of the researchers. This computer system must have a user interface, a method of data validation and a persistence layer. The user interface is responsible for entering data into the system, the validation method is responsible for validating the data entered so that relevant data are not forgotten or recorded incorrectly and the persistence layer is responsible for storing the data in the database mobile devices. The persistence layer must necessarily store the data in a DBMS that supports spatial data. After collecting

the data, there is a synchronization step, where the collected data is synchronized with the synchronization server. The objective of synchronization [3] is to generate replicas, through a full replication, of the collected information to the server that synchronizes data with the database server involved in the core collection. This synchronization is specified in the next section. Finally, there is the storage of the data (via a data backup) servers in the core data set involved in the collection. The possibility of having more than one server central database is taken into consideration in this architecture, given the possible dimensions of data collection, which allows the flexibility aspect to the architecture. When we have more than one server database replication, server data synchronization occurs for these servers, creating data redundancy, which prevents the architecture from having a single point of failure. The synchronization protocol is used in architecture that will lay down the interoperability and flexibility of the architecture (flexibility towards the possibility of change, to adapt to contexts of different data collection), so the protocol for the architecture should be characterized as interoperability, and should be a free protocol.

Synchronization Stage

The proposed synchronization is divided into two stages: capturing SQL statements and the transfer of these instructions to the synchronization server. The capture module modifications are responsible for capturing the SQL database from the mobile device in the event of any change in the database. The transfer module modification plays a client role in a client-server architecture, synchronizing the changes to the SQL statement capture module with the synchronization server.

Capture Module Modifications

The capture of these new SQL statements is by searching for data that have a flag with value 1. This flag determines if the data has been synchronized or not. If it is set as 1, then the data has not been synchronized yet; if it is set as 0, then the data has been synchronized. This search can occur in all tables of the database or only one table in the database that concentrates all the changes that have occurred (See Figure 3).

Transfer Modifications Module

After the capture of new instructions, the modifications transfer module makes the transfer of these instructions to the server synchronization [13]. This transfer occurs according to the synchronization protocol chosen for the architecture, which, as previously mentioned, should be a protocol that has the characteristic of interoperability and is free (Figure 3).

The flow diagram of these modules shows if a connection error occurs when transferring changes, the variable n (the variable in the conditional structure of the flow diagram) will have the value of the SQL statements that were not transferred. Thus, when the internet connection is reinstated the transfer continues from where it stopped without causing data inconsistency.

CONCLUSION:

The existing data collection architectures satisfy a specific cause, using specific technologies that are limited regarding the type of data, type of networks, synchronization type, device type, and so on. For each different data collection context architecture is proposed for data collection with different budget limitations, computational environments and available technology within the company or organization. The proposed architecture for this work was an architecture of data collection for field research in heterogeneous computational environments that supports vector geographic data where information stored on each mobile device are replicated and correctly integrated into a central database. To reach this goal the architecture had to obey two aspects: interoperability and flexibility. This architecture complies with the aspect of interoperability, using SyncML synchronization protocol, and the aspect of flexibility, by replicating data amount where there is an intermediate server and a free synchronization protocol. Also, the architecture uses some safety mechanisms such as fault tolerance, full replication data and authentication. With this architecture, it is no longer necessary to plan architecture whenever you make a data collection in the field, just use the proposed architecture, which in addition to functioning in heterogeneous environments and support vector geographic data, is low cost and high quality.

REFERENCES:

[1] https://www.fhi360.org/sites/default/files/media/documents/Paper_to_Mobile_DataCollection_Manual_1.0.pdf
[2] http://blogs.worldbank.org/edutech/using-mobile-phones-data-collection-opportunities-issues-and-challenges.
[3] https://www.academia.edu/6590765/Data_Collection_Architecture_for_Field_Research_in_Heterogeneous_Computational_Environments
[4] https://www.globalquakemodel.org/media/publication/DATA-CAPTURE-GEM-Userguide-Android-Mobile-Tool-201403-V01.pdf.