Mobile Assisted Personal Interview Software-Development and Experiences from Implementation in Crop Estimation Survey in India

Kaustav Aditya[1], Hukum Chandra[1], Sushil Kumar[1], Niranjan Nayak[1] and Shrila Das[2]

[1]*ICAR-Indian Agricultural Statistics Research Institute, New Delhi*
[2]*ICAR-Indian Agricultural Research Institute, New Delhi*

SUMMARY

The use of conventional method of survey data collection, i.e. Paper Assisted Personal Interviewing (PAPI) method has several inherent weaknesses such as non-sampling errors in collection, tabulation and processing of the data. This leads to problem in data quality and timeliness. This article describes an android based Mobile Assisted Personal Interview (MAPI) software for collection of survey data using smart phones. The MAPI software developed by the ICAR-Indian Agricultural Statistics Research Institute (ICAR-IASRI), New Delhi, has been implemented successfully in the agricultural survey conducted in the two states in India namely Uttar Pradesh and Gujarat. The results obtained from these surveys show that MAPI is efficient both in time and accuracy and emerges as an efficient alternative to PAPI.

Keywords: Android smart phones, Survey data, MAPI, PAPI, Non-sampling error, Census.

1. INTRODUCTION

Objective data collection, analysis and presentation serve as basis for taking policy decisions in different fields of daily life. The important users of statistical data, includes Government, Industry, Business, Research Institution, Public and International Organizations. The inferences drawn from the data help in determining future needs of the nation and also in tackling social and economic problems of people. The industry standard conventional approach of paper based data collection, i.e. PAPI method, requires huge amount of resources including man power, cost and time. Due to shortage of man power, increasing activities and workload, survey implementing agencies are under tremendous pressure and hence quality and timeliness of data becomes questionable. This also leads to occurrence of various non-sampling errors in the survey data, at various stage including data entry and data processing. In the spirit of Computer Assisted Personal Interviewing (CAPI) (Wikipedia, 2016), the MAPI software developed by the ICAR-IASRI, New Delhi, helps in collecting the

quality and timely data. This software is developed based on android platform as more than 90% of Indian smart phone market is occupied by Android based smart phones. Both online and offline version of the MAPI software has been developed. The software has been copy righted under Indian copy right act with reg. no. SW-9378/2017. The MAPI has been tested and implemented in two states in India namely Uttar Pradesh and Gujarat under a pilot survey entitled "Pilot study for developing State level estimates of crop area and production on the basis of sample sizes recommended by Professor Vaidyanathan Committee report" conducted by the ICAR-IASRI, New Delhi. In particular, the MAPI has been implemented in the Bulandsahar and Pratapgarh district of Uttar Pradesh during Rabi season of Agricultural year 2015-16 and in Gandhinagar district of Gujarat during kharif and Rabi season of Agricultural year 2016-17 in India. Three questionnaires were prepared and implemented using for both MAPI and PAPI to collect data on crop area and yield statistics from the selected villages of each

Corresponding author: Kaustav Aditya
E-mail address: katu4493@gmail.com

district of the two states under the pilot project. The rest of the paper is organized as follows. Next Section discusses about the development of the software and its process flow diagram. Section 3, describes the field testing of the MAPI software, Survey design and data collected. Section 4, enlightens the outcome of the field testing of the MAPI software against PAPI as well as summary of the study and Section 5, consist the conclusion of this study.

2. MOBILE ASSISTED PERSONAL INTERVIEW SOFTWARE

2.1 Steps for using MAPI (offline version) software:

The MAPI Android Application has been developed using the Eclipse IDE (Pandey and Dani, 2014), which is a suite of programs that facilitate developing and running codes written in the Java (Hanna, 2003). In MAPI the inbuilt Microsoft SQL Server (Date *et al.*, 2006 and Nielsen and Parui, 2011) under Eclipse IDE was used for creating the data base while collecting the survey data. The software was developed for android version 4.1 (Jelly bean) and is compatible with all available android smart phones in the market. To develop MAPI, android eclipse software has generated .apk (android package kit) file for installation in the mobile devices. For agricultural survey data collection, questionnaires have been developed as per the requirement of the pilot project for PAPI and then it was customized for MAPI for collection of data using smart phones. The software is available online at Sample Survey Resource Server of ICAR-IASRI, New Delhi at http://sample.iasri.res.in/ssrs/android.html. For other surveys the developed questionnaires in the software can also customized based on the requirements of the registered user/organization. Recently, various other states in India have also shown interest in implementing MAPI in their ongoing surveys as an alternative to PAPI. The process flow diagram of MAPI software is shown in Fig. 1 for offline version.

MAPI android App. works in three Steps:

1. User Interface
2. Schedules/Questionnaires (Three questionnaires are made for the survey, i.e. Enumeration schedule for collection of data on cropped area and CCE schedule 1 and 2 for collection of data on crop yield.)

3. Storing of data in Database of the device and generation of the Ms-Excel sheet.
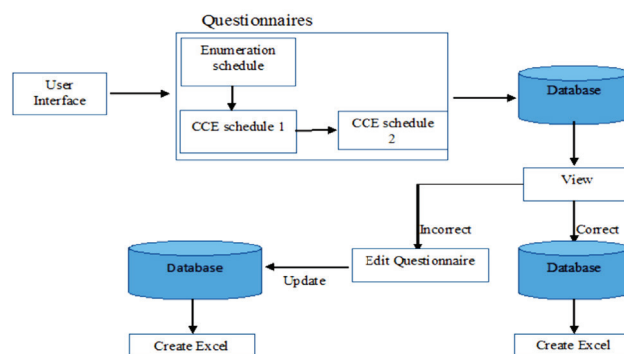

Fig. 1. Process Flow Diagram of MAPI

2.1.1. User Interface

Steps to access MAPI in android support device,

Download the software from http://sample.iasri.res.in/ssrs/android.html of the Sample Survey Resource Server of ICAR-IASRI, New Delhi.

1. Install the setup file named "MAPI.apk" in memory of the smart phone or tablet.

2. After installation, MAPI app icon displayed on the main screen of the device. The "Menu" tab in MAPI is provided for the users for navigation between several questionnaires at the time of survey.

3. As shown in the home screen of the application, user can easily visit ICAR-IASRI website with simple click. However to use the questionnaires, user has to signup first with the application as shown in Fig. 2.

2.1.2. Questionnaires

After filling one time registration form users need to click on create button which entitles them for login to the application and access the questionnaires (Fig. 2). The software has been built with a remember password tab which helps the users in frequent login to the software without any constraint. The software has been made to fit with the questionnaires prepared under the pilot survey planned at the institute for enumeration of crop area and yield under each selected village of each selected tehsil of each district of the two selected states of India.

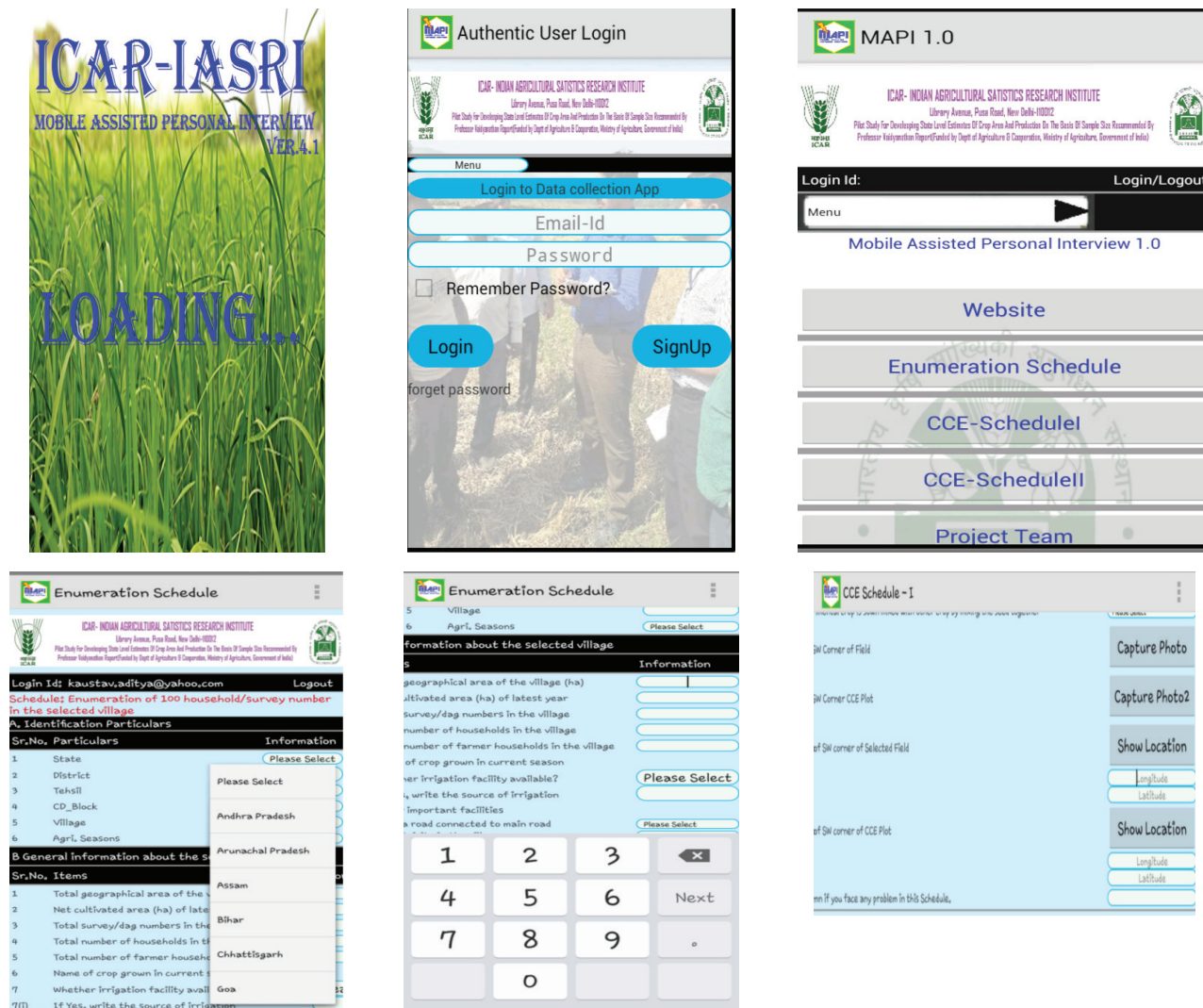The following questionnaires are used in the MAPI software.

Fig. 2. User interface and Questionnaires of MAPI

a. Enumeration Schedule for Crop Area Enumeration

The Enumeration schedule (see Fig. 2) was formed to collect data on cropped area, irrigation status, crop under cultivation, farmer's name, farmers survey number, area of the farming household (in Hectare), name and number of crops cultivated etc. In this questionnaire we collect data to create a sampling frame out of each selected village for selection of plots/parcels/survey numbers make for conduct of Crop Cutting Experiments (CCE) (Sud *et al.*, 2015) for crop yield rate enumeration.

b. CCE Schedule-I: For selection of plots for CCE

Particulars of plots/parcels selected for crop cutting experiment is recorded in CCE schedule-I.

CCE Schedule-I consists of questions to record the GPS and pictorial information of the selected plot of CCE as shown in Fig. 2. These questions were added for verification of the user or field investigators that actual visit to the plot has been made to ensure the data validity. The users have to complete the whole questionnaire before submitting the data using the "Submit" button in the questionnaire.

c. CCE Schedule-II: For recording the produce obtained from CCE plot

This questionnaire is formulated to capture the information regarding the outcome of CCE from the farmer's field. It records the amount of produce obtained from the CCE plot in Kilogram (Kg) with date of harvesting of the crop.

2.1.3. Data Storage and Creation of the MS-Excel sheet

After filling up the questionnaires the data gets stored in the MySQL database of the device. To create and access the MS-Excel file, the user has to view the data stored in the database by clicking the "View" button. After viewing the data if the entries are correct then the MS-Excel file can be generated. If there is any mistake performed during the process of data collection the user has to edit the data by clicking the 'Update' button (see Figure 10).

2.1.4 Steps for using MAPI (online version) software:

In MAPI (online version) the Microsoft SQL Server 2008 was used for creating the data base while collecting the real time survey data. Net beans IDE have also been used for creating the web services which connects Eclipse IDE with SQL server during the process of data entry, updation or deletion. The first two steps of MAPI online and offline version are same while in the third step data is directly synchronized with server located at the headquarters through web service.

3. FIELD TESTING OF MAPI SOFTWARE

For field testing, the software has been implemented during the Agriculture Year 2015-16 in two districts namely Bulandsahar and Pratapgarh of the state of Uttar Pradesh in India and in Agriculture Year 2016-17 in Gandhinagar district of Gujarat state of India. In Bulandsahar and Pratapgarh, one tehsil/sub-division from both the districts was considered for Rabi season of Agricultural Year 2015-16. The selected tehsil/sub-division from Bulandsahar and Pratapgarh districts were Bulandsahar and Kunda respectively. From the selected tehsils a total of 12 villages and 8 villages were chosen respectively. The list of selected villages was given in Table 1 and 2. From each village, information about crop area and yield of the major food grain crops prevailing in that region was collected using the questionnaire developed under the pilot project for both MAPI and PAPI. Due to unavailability of proper cellular network in the selected villages under this study we are only be able to use the offline version of MAPI software. The most predominant crop during the season was wheat while few pulses along with few other oilseed crops mostly rapeseed and mustard were found in those selected villages. Almost 90% area was found occupied by wheat crop in those villages. In case of Gandhinagar district of Gujarat two tehsils

Table 1. Mean, Sum and Standard Deviation (SD) of Area (in Hectare) for Paddy based on survey by PAPI and MAPI in Pratapgarh district of Uttar Pradesh

Village	PAPI			MAPI		
	Mean	Sum	SD	Mean	Sum	SD
Bhulsa	0.27	12.58	0.23	0.26	11.26	0.28
Chakethi	0.39	29.77	0.45	0.38	16.53	0.47
Kalyanpur	0.21	05.41	0.15	0.21	04.32	0.16
Purechiranjeev	0.43	13.26	0.26	0.42	13.32	0.26
Salhepur	0.38	17.64	0.91	0.39	17.20	0.93
Samapur	0.19	05.19	0.32	0.18	04.95	0.32
Shivpur	0.45	12.19	0.25	0.41	10.98	0.32
Ebrahimpur				0.23	7.22	0.25

Table 2. Mean, Sum and Standard Deviation (SD) of Area (in Hectare) for Paddy based on survey by PAPI and MAPI in Bulandsahar district of Uttar Pradesh

Village	PAPI			MAPI		
	Mean	Sum	SD	Mean	Sum	SD
Akbarpur Raina	19.80	39.60	1.56	19.80	39.60	1.56
Auledha				19.35	38.70	0.92
Dhamrawali	18.40	36.40	2.82	18.55	37.10	3.04
Faizabad Rithavali	15.80	31.60	1.13	15.78	31.55	1.17
Khwajpur Asaraf	18.90	37.80	0.56	18.90	37.80	0.57
Lahagra	18.40	36.80	3.82	18.38	36.76	3.85
Meerpur	14.05	28.10	9.69	14.05	28.10	9.69
Pouth	19.00	38.00	0.71	19.00	38.00	0.71
Rahimpur Behegavan	16.90	33.80	0.14	16.75	33.50	0.35
Rajgarhi	16.45	32.90	2.47	16.48	32.95	2.51
Salvat Nagar Gangawali	18.85	37.70	0.21	18.83	37.65	0.25
Tatarpur	15.80	31.60	1.84	15.78	31.55	1.87

namely Dehgram and Mansa were selected. A total of 12 villages were selected in Gandhinagar where from 7 villages from Dehgram tehsil and 5 villages from Mansa tehsil for Agricultural Year 2016-17. Under the pilot survey only major food grain crops are covered. In Gandhinagar four prevalent crops are found in the selected villages namely, paddy, wheat, Black Gram (Urad) and Pearl millet (Bajra).

Under the pilot survey, first selected 100 survey numbers were selected randomly from the frame of total survey numbers in each village in clusters of 5 survey numbers using equal probability without replacement sampling design. For each survey number visits were made to the farmer who was associated with

that survey number and information were collected about total cropped area, net cultivated area and crop cultivated through the Enumeration Schedule. This work was done within one month of sowing of the crop. Then 15 days before harvesting of the crop, two survey numbers are randomly selected out of selected 100 survey numbers and from the selected two survey numbers information was collected from the CCE plot using CCE schedule I and II respectively. To implement the MAPI software Samsung galaxy Tab 4 Tablet loaded with android ver.4.4 was used. To verify the functionality of MAPI software, the data was compared with the data collected through traditional PAPI survey in the selected villages under the pilot project with the help of the state officials of Department of Agriculture and Crop Insurance, Lucknow, Govt. of Uttar Pradesh and Agro Economic Research Centre (AERC), Anand, Gujarat of India. In the following section we discuss the results from the comparison of MAPI vs. PAPI.

4. RESULTS AND DISCUSSION

The following are the results of the comparison study of the data that was obtained from MAPI and PAPI after data entry was done by the state officials. Table 1 and 2 contains the summary of area data in hectare (ha) for paddy crop collected through PAPI and MAPI in Pratapgarh and Bulandsahar district respectively of Uttar Pradesh state during Rabi season 2015-16 and Table 3 and 4 contains the summary of the data from Gandhinagar district of Gujarat state for paddy crop in Kharif season 2016-17 and wheat crop in Rabi season 2016-17. From Tables 1 and 2, it is clearly visible that for Ebrahimpur and Auledha villages of Uttar Pradesh the whole data is missing for PAPI survey. Further, from Table 3 and 4 it is evident that there is a huge amount of dissimilarity between the PAPI and MAPI data as in table 3 the data for the entire tehsil Mansa is missing for area on paddy crop. From table 4 it is evident that the data obtained from MAPI and PAPI is completely different. The main reason behind these results was the mistakes committed by the data entry operators at the time of data entry at the state headquarters. The error occurred due to this kind of mistakes are beyond control of any organizations involved in survey which on the contrary reduces the sample size and increase in standard error of the estimates.

The empirical results reported in Table 1-4 indicate that in some cases the SD of the data collected through MAPI is higher than the PAPI. A critical investigation of data and empirical results reveal that there are a lot of repeated values of observation in the data reported through PAPI. These repetitions have occurred in the form of same farmer or survey number or same value of observation for many survey numbers in a village. This clearly an example of non-sampling error in PAPI based data collection. It is difficult to rectify such non-sampling error at the estimation stage when data has been collected and tabulated by some other agency. This investigation of results also highlight an important conclusion that the smaller value of SD for the PAPI based data than the MAPI does not mean that data quality is good from the PAPI. The MAPI data is free from non-sampling error because of its in-built checked and conditions, and hence leads to more realistic results.

One of the major concerns of development of MAPI is timeliness. In our country, the process of crop estimation survey takes a total of 5 years in which the first three years were spent only for data collection and data entry. To ease this process and reduce the time wasted during the above mentioned process, MAPI software was developed with funding from Directorate of Economics & Statistics, Department of Agriculture, Cooperation and Farmers Welfare, Ministry of Agriculture and Farmers Welfare, Govt. of India under the pilot project. In case of MAPI, data was received instantly survey as soon as the data collection is over in the field whereas in case of PAPI the traditional method is followed for data collection and data entry. Hence, after data entry, the data were received during April 2017 which is around 12 months after completion of the survey in case of Uttar Pradesh state of India while in case of Gandhinagar district of Gujarat PAPI data was received in March 2018 almost 11 months after completion of the survey as compared to MAPI data which was received during May, 2017. This kind of problems prevails in almost all the states and developing countries where regular surveys are conducted and due to these problems the outcomes of the surveys get delayed further delaying the policy decisions forthcoming out of the survey. Hence, the fact that MAPI is a faster and easier way of data collection than the traditional PAPI surveys can be easily established. Further, MAPI software with inbuilt checks and conditions reduce certain non-sampling errors like, name of state, block, tehsil, villages, crops etc. as they are kept as prefilled and only be selected from the drop down list of names, also there were checks for area of

Table 3. Mean, Sum and Standard Deviation (SD) of Area (in Hectare) for Paddy based on survey by
PAPI and MAPI in Gandhinagar district of Gujarat for Kharif season 2016-17

Tehsil	Village	PAPI			MAPI		
		Mean	Sum	SD	Mean	Sum	SD
Dehgam	khanpur	-	-	-	04.45	08.90	0.21
Dehgam	Ahamadpur	0.61	01.22	0.43	0.61	01.22	0.43
Dehgam	Amrajina Muvada	0.25	01.26	0.37	0.25	01.26	0.38
Dehgam	Bardoli (Bariya)	0.78	07.06	01.07	0.78	07.06	01.07
Dehgam	Dhamji	0.66	23.01	0.82	0.66	23.01	0.82
Dehgam	Sagdalpur	0.78	22.48	0.73	0.78	22.48	0.73
Mansa	Lakroda	-	-	-	03.05	06.11	0.07
Mansa	pratapnagar	-	-	-	02.91	05.85	0.14
Mansa	Vijaynagar	-	-	-	04.61	09.23	0.14

Table 4. Mean, Sum and Standard Deviation (SD) of Area (in Hectare) for Wheat based on survey by
PAPI and MAPI in Gandhinagar district of Gujarat for Rabi season 2016-17

Tehsil	Village	PAPI			MAPI		
		Mean	Sum	SD	Mean	Sum	SD
Dehgam	Ahamadpur	05.73	11.46	02.82	05.73	11.46	02.82
Dehgam	Amrajinamuvada	02.81	05.61	03.11	02.81	05.61	03.11
Dehgam	Anguthala	04.65	09.29	0.67	04.65	09.29	0.67
Dehgam	Bardoli	06.21	12.50	01.09	06.25	12.50	01.06
Dehgam	Dhamij	06.85	13.65	0.45	06.85	13.65	0.45
Dehgam	Khanpur	07.19	14.38	03.10	07.19	14.38	03.10
Dehgam	Sagdalpur	07.00	14.00	01.05	07.00	14.00	01.02
Mansa	Bapupura	08.46	10.78	0.49	08.46	10.78	0.49
Mansa	Lakroda	09.30	18.76	01.44	09.38	18.76	01.47
Mansa	Pratapnagar	07.05	14.11	0.21	07.05	14.11	0.21
Mansa	Rajpura	08.50	16.99	0.54	08.50	16.99	0.54
Mansa	Vijaynagar	08.68	17.03	0.57	08.52	17.03	0.55

the plot or field which should always be less than total geographical area of the village and many more. Also, there was always a chance for occurrence of error by the contractual data entry staff during the process of data entry. Also if the question is, "What is the area under a crop?" and the answer to be obtained should be certain numerical value with fixed unit of area i.e. hectare (ha) or Acre. But, after receiving the data from PAPI, the unit of area and yield of the villages became the major problem while analyzing the data as some enumerator have put the area in hectare, some in acre and some in other local units and same in case of weight of produce. Due to these problems, the data obtained from PAPI was processed by extensive cleaning and scrutinizing. In case of national level surveys i.e. the General Crop Estimation Surveys (GCES) conducted by Directorate of Economics & Statistics, Department of Agriculture, Cooperation and Farmers Welfare, Ministry of Agriculture and Farmers Welfare, Govt. of India, the bulk of data is huge and with these errors of PAPI it is almost impossible to scrutinize and clean the data in due time which leads to delay in generation of crop statistics. To solve this kind of problems in a single go, in MAPI, the area unit is kept fixed with hectare (ha) as standard and weight as Kilogram (kg). In MAPI, if needed the enumerator have to convert the local unit in the standard unit before giving the input. These advantages, checks and restrictions of MAPI eliminate several errors usually appears during large scale surveys making it a more reliable and efficient substitute of PAPI.

Another goal of MAPI is to validate the data i.e. it was actually collected from the field where the survey is planned. It is observed from the past surveys that many a time, surveyors fill the data without visiting the field and with their own intuition due various reasons. For this purpose questionnaires are included a in the MAPI regarding recording the GPS location of the field where the actual survey is taking place. The enumerator have to record the GPS location of the field where the survey is taking place along with the location of the south west corner of the CCE plot used for crop yield estimation. Along with the GPS location there is also a questionnaire regarding taking the photo of the field under the survey along with the CCE plot to verify that the enumerator actually visited the field. Further, there was a hidden questionnaire that records the date and time of collection of the data along with the email id of the field investigator which gets stored automatically in the data and it was visible when the data was accessed. As an example, a field enumerator collected the data from Gandhinagar district of Gujarat as proposed where as from the GPS location of the survey is recorded as Longitude: 72.9308867 and Latitude: 22.5550971. After searching the location online to verify that is it the same location where the survey needs to be conducted. It was found that the enumerator has collected the data from Anand district of Gujarat state in India which is almost 100 kilo meters away from Gandhinagar. The other coordinates point at the exact location where the survey is to be conducted and the data is collected by a separate field enumerator. This kind of situations is very common in surveys and due to these problems questions always arise about the reliability of the data under analysis. The study reveals that the MAPI software performs much better than the traditional ways of data collection in large scale surveys i.e. PAPI in India from the point of view of timeliness, validity and reliability of the data which is a major national concern.

5. CONCLUSION

Software is generally made to reduce the quantum of work. Based on the results obtained from a real time pilot survey we can conclude that the MAPI software will reduce the paper work and increase reliability of the collected data. With inbuilt checks and conditions for validity of data, the data received from MAPI software is easy to process and ready to use for analysis purposes saving lot of time, resource, man power and money, making it a better way of collection of data based on the current scenario of use of smart phones and tablets in our regular life. The successful implementation of MAPI in two pilot states also motivated other states in India to use this software for data collection in all surveys conducted by them. This software will help the policy planners and other Government organizations involved in conducting surveys for determining future policies in collection of data in fast and efficient way without wasting a lot of time. Further, the software is cheap and easy to employ as compared to its other alternatives currently available in the market. It also has an advantage over other softwares currently available in the market as those softwares requires intensive training and specific set of tools developed by their own to be installed in the secured server where all the data is going to be stored which does not ensure complete security of the data collected where as MAPI is independent of all those problems as the data gets stored in the server of the user and only the user can access the data without any interference of software developers/organizations.

REFERENCES

Date, C.J., Kannan, A., Swamynathan, S. (2006). An Introduction to Database Systems. Pearson Publication.

Hanna, P. (2003). The Complete Reference JSP 2.0. Tata McGraw Hill Education Private Limited, New Delhi.

Nielsen, P. and Parui, U., (2011). Microsoft SQL server 2008 bible (Vol. 607). John Wiley & Sons.

Pandey, G. and Dani, D. (2014). Android Mobile Application Build on Eclipse, International Journal of Scientific and Research Publications, Volume 4, Issue 2, February 2014.

Singh, R. and Mangat, N.P.S. (1996). Elements of Survey Sampling. Kluwer Academic publishers.

Sud, U.C., Ahmad, T., Gupta, V.K., Chandra, H., Sahoo, P.M., Aditya, K., Singh, M. and Biswas, A. (2015). Gap Analysis and Proposed Methodologies for Estimation of Crop Area and Crop Yield under Mixed and Continuous Cropping under the project "Research on Improving Methods for Estimating Crop Area, Yield and Production under Mixed, Repeated and Continuous Cropping". Working Paper No. 4. Global Strategy, FAO, Rome Publication. Available at: http://gsars.org/wp-content/uploads/2015/12/WP-2-on-Improving-Methods-for-Estimation-of-Crop-Area-and-Crop-yield-under-Mixed-and-Continuous-Cropping-141215.pdf.

Wikipedia, the Free Encyclopedia, 2016. Computer assisted personal interviewing (CAPI). Available at https://en.wikipedia.org/wiki/Computer-assisted_personal_interviewing.