A digital approach to surveys in developing countries

11 authors, including:

Some of the authors of this publication are also working on these related projects:

Volunteerism and Youth Employment View project

A digital approach to surveys in developing countries

MILLE Guillaume, TOUCH Franck
KhmerDev, #184, St Monireth (217), Phnom Penh, Cambodia

BARAN Eric, NASIELSKI Joshua, SARAY Samadee, TAN Sokhom, TRESS John
WorldFish, P.O. box 1135, Phnom Penh, Cambodia

CHEACH Monyda, KHUN Seila, MEN Pagnchak Roat, SREY Sinath
Cambodian Agricultural Research and Development Institute, National Road No. 3, Phnom Penh, Cambodia

Abstract

We present here a methodology developed for gathering socioeconomic and environmental survey data in the field. The system is divided into two interfaces: i) an Android tablet with survey software for data collection in the field and ii) a back office application for questionnaire management and data compilation. Each of these technical components is detailed and the solutions to challenges and constraints that were faced are described. This system allows the survey software to be updated in real time; data is collected daily from multiple tablets through an ftp server and combined into a single matrix ready for analysis. The key strength of this system is its user friendly interface, its time- and cost-efficiency and its road-tested reliability in developing countries, even in remote places.
Keywords: Android, tablet, back office, socioeconomics, environment

1. Introduction

Development interventions in developing countries are often guided by field-based surveys aimed at assessing socioeconomic or environmental status in order to identify issues, prioritize regions and target communities or actions [1]. Until recently, surveys were implemented using paper questionnaires in the field, and data was subsequently entered into computer databases by hand. This process is heavily time consuming, costly and subject to multiple errors posterior to the initial data recording [2].

The recent development of digital devices is changing the conditions of field surveys. The marketing sector has been at the forefront of such development, but in research it is mainly the health sector that pioneered the use of portable digital devices for data gathering, starting with digital assistants or pen-tablets [3] and handheld computers [4]. In the following years field-based collection of data with digital devices was further conceptualized [5] and applied to an expanding range of disciplines [6]. In socioeconomics the Asian Development Bank has recently promoted the use of Tablet Computers for Household Surveys in Asia, since it saves "time and money, while improving data integrity" [7].

We present below the tablet-based application developed for socioeconomic surveys in rural Cambodia. After one year of use, the approach has proven robust, effective and cost effective even in remote locations characterized by limited infrastructure. Its versatility in a context of extensive and rapidly expanding cell phone network coverage makes it a tool of choice for surveys in developing countries.

2.   Material and Methods

A questionnaire covering socioeconomic, health and environmental aspects was developed to quantify the multiple values of fish resources in Cambodia and assess the dependency of rural households on this natural resource (*www.facebook.com/MekongFishValue*). This 265 questions long questionnaire was used in 747 households randomly selected throughout the country, using 10 digital tablets for data recording. The questionnaire was developed in MS Excel format before being coded for its integration to an Android environment. The tablet-based survey application developed consists of 2 main components: a tablet interface for collecting data and a web-based back office application to manage the questionnaire and the collected data.

2.1. The tablet application

This application is designed for 10 inch screen Android tablets. The main challenge was the versatility required, in particular the need of:
- a database model able to handle any type of question in any order;
- a display able to handle those similar constraints as well.
- an option to remotely update questions or add new ones.
The first step consisted in designing a database model that would not require an update every time a new type of question had to be handled [8]. The model created is illustrated in Figure 1.
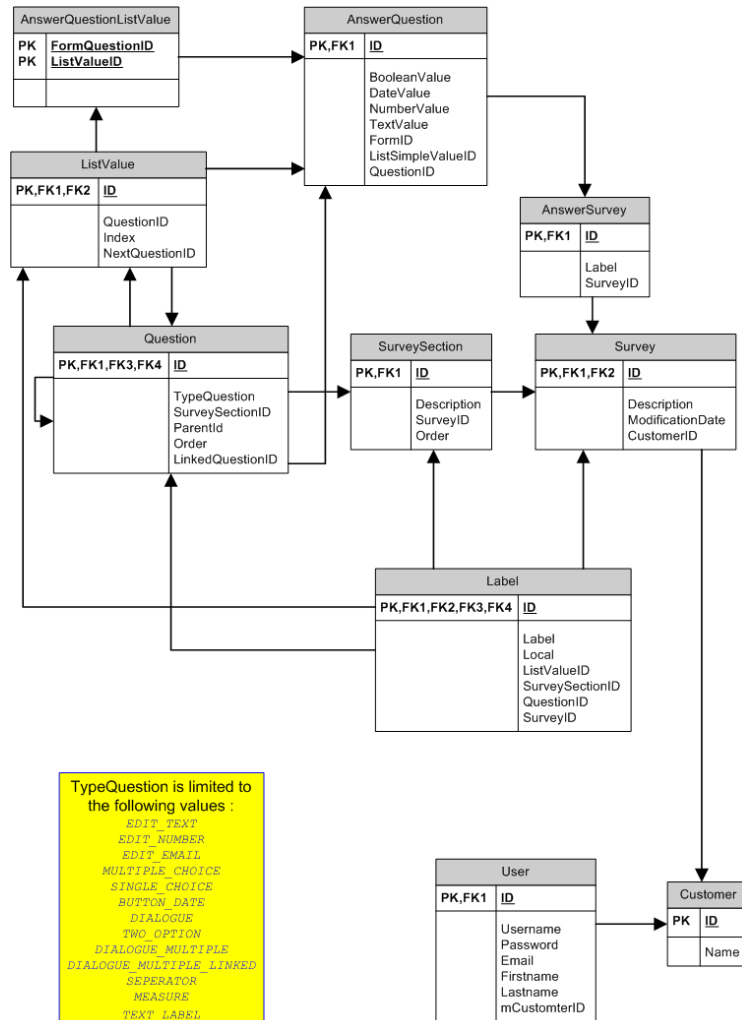
Figure.1 Database model of the questionnaire tablet application.

The key components of the model are tables Question and AnswerQuestion. Table Question stores all of the questions asked; the content of this table shapes the questionnaire displayed. This table includes some specific fields, in particular:
- TypeQuestion: the type of question, which will determine how the question will be displayed in the application and how it will be stored.
- ParentID: when questions are followed by sub questions, each sub-question is referenced to a Question through this field. A sub-question can itself be the parent of another question and there is no limit to such series, although long series alter performance and usability.
LinkedQuestionID: a type of question called "DIALOGUE_MULTIPLE" allows asking several times the same subset of questions. Then the answers to the subset

of questions must be related together using the LinkedQuestionID field (for instance questions about the costs and benefits from various activities are related to a person).

Table AnswerQuestions stores all the answers collected. This table includes some specific fields, in particular:

- QuestionID: identification of the question answered;
- AnswerSurveyID: links a given answer to an AnswerSurvey, which is the entity that puts together a set of answers for a given survey or questionnaire;
- ParentAnswerID: field used for "DIALOGUE_MULTIPLE" questions; when the same subset of questions is answered many times, each subset of answers must be linked to a parent Answer.

As for the user interface, the biggest challenge faced was the dynamic nature of the questionnaire, and the need for the surveyor to be able to flip ahead in the survey or come back to previous questions. We used Fragments, i.e. Android components that allow managing only dynamism in the code, while the presentation is managed through some static xml layouts [9].

We created an empty shell in which fragments are added/removed according to the questions defined in the database. Each type of question has a corresponding fragment and a new fragment instance is created for each question according to its type. For this, an implementation of the strategy design pattern is used [10]. Moreover, the fragments were designed so that they know how to store the values inputted into them. Figure 2 shows that each Fragment can initialize itself, save its value, and which type of question it supports [11]. Then FragmentManager checks whether each Fragment supports the type of Question, then upon positive response it initiates the implementation.
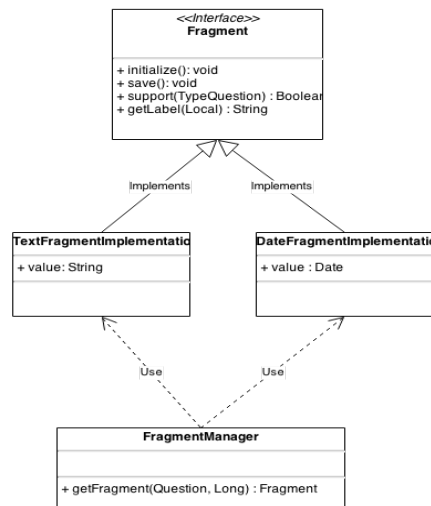

Figure 2.Simplified class diagram for the classes involved in the management of fragments. Only 2 Fragment implementations among 14 are shown here.

2.2. The Back Office

Thus, the Android tablet application displays questions as they are defined in the database and stores the answers. Beyond that, users want to be able to retrieve answers and analyze them; we designed for this purpose a user-friendly database model for data restitution towards a MS Excel file.

One challenge arose from the DIALOGUE_MULTIPLE type of questions, since the final number of answers is not a priori known, and yet a standardized data reception matrix of fixed dimensions is required for data analysis. Moreover it is not possible to fix a number of columns matching the maximal number of answers collected in a given tablet, since this number varies from tablet to tablet and each device is not connected to the others. We circumvented the problem by adding a maximum number of answers for each question with multiple answers, expanding the matrix to dimensions superior to the possible number of answers [12].

Additional requirements were i) the possibility to update questions in the questionnaire, without having to manually update the database in each tablet, and ii) the need to automatically compile the files (i.e. the individual questionnaires) exported from each tablet.

A web-based back office was developed to meet the demand .This back office allows adding and removing questions, updating their labels, changing their order and creating new questionnaires; it also compiles all the individual files received from all the tablets into one single data matrix.

The database model of the back office is the same as that of the tablet application, plus a few tables to handle the user interface, in order to make communication between the back office and the tablet application easier.

A synchronization mechanism allows this back office to communicate with the application on each tablet; for that purpose an asynchronous communication protocol was implemented as described in Figures 3 and 4. Thus, a ftp server centralizes every update from either the tablet application or the back office.
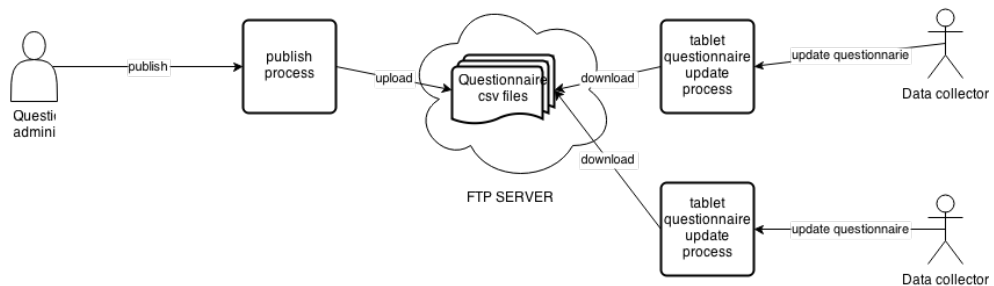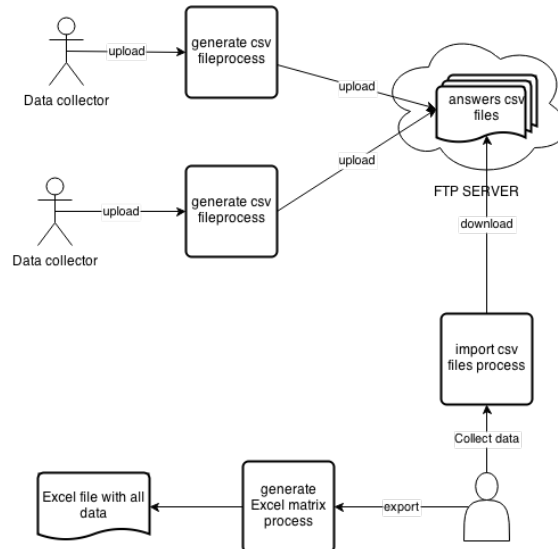

Figure 3.Synchronization: updating a questionnaire.

Figure 4.Synchronization: gathering the collected data to the back office and exporting them.

Questionnaire updates are done through the back office; the changes are stored only in the back office without being immediately sent to tablets due to the risks of spamming and to allow double-checking updates before validation: the user publishes the update only when all changes are validated. It is during this publishing action that the tables describing the questions of the survey (Survey, SurveySection, Question, Label, ListValue) are exported in csv format. The csv format was chosen because it is widely used and easy to write and parse. Then the tablet users update the survey, which is straightforward since the data model is the same and no mapping is required: the existing questionnaires are simply replaced with the latest ones.

Data compilation starts when the user of the tablet application triggers the export of the tables in which data are stored (AnswerSurvey, AnswerQuestion and AnswerQuestionListValue); these data are exported in csv format and uploaded to an ftp server. The back office downloads these files from the ftp server and imports them in the back office database of the same data structure. The point is to prevent the same data from being imported twice. This is ensured by not importing the questionnaires whose identifiers (stored in the AnswerSurvey table) are already in the database. Since Android devices have unique identifiers, the latter can be added to the database identifiers to get truly unique identifiers for each questionnaire.

Converting multiple large csv data sets such as those gathered in multiple tablets during the survey into a single Excel file matrix was challenging (3,500,000 Excel cell and more than 1 hour required for compilation), so we created a planned task running every night and generating a file reflecting the latest data

compiled. The drawback is only that the end user must wait one night after the last survey to get the whole data set ready for analysis.

3. Results

Thanks to the methodology described above, the project "Assessing economic and welfare values of fish in the Lower Mekong Basin" could gather answers to 197,955 questions regarding socioeconomic, health and environment issues in 37 villages scattered throughout the country, most of these villages being isolated and having no access to electricity. Data gathered during the day were sent in the evening, using the telephone coverage of the village surveyed or of a nearby village with a telephone connection. Updates of the questionnaire could be downloaded within 24 hours in the tablets even in remote locations. The standard battery capacity of the Samsung tablets allowed using them day long for questionnaires, provided that Saving mode and Flight mode (i.e. no network activity) were activated. The tablets were recharged each night using 220V when available, or alternatively 12V car batteries common in remote villages. Thanks to sturdy cases, none of the 10 tablets was damaged or failed after one year despite the rough conditions of use. One survey round (747 households), implemented using ten digital tablets, took three months and provided data ready for analysis the day after the last questionnaire was filled. The final Excel format chosen also allowed partners unfamiliar with databases to undertake analyses.

4. Conclusions

This research detailed here shows that the technology developed using digital tablets for surveys in tropical environment has proven effective and reliable. We are currently using the back office to create a new questionnaire on fish migrations, which illustrates the fact that the tool is versatile and not bound to a given type of questionnaire or discipline. Among perspectives, on the technological side, are new features such as conditional branching in questions (choosing the next question according to the previous answer), adding constraints (like min and max for a given answer), or adding integrated reporting to the back office. As already recognized by the Asian Development Bank, digital tablets provide new opportunities for surveys in developing countries, and this avenue deserves attention and investment.

References

[1] Nossin J. J. *Surveys for development: a multidisciplinary approach*. Elsevier Scientific Pub. 1977.

[2] Marsh R., Drowning in dirty data? It's time to sink or swim: a four-stage methodology for total data quality management. *Journal of Database Marketing and Customer Strategy Management*. 12 (2):105-112. 2005.

[3] Spain K.A., Phipps C.A., Rogers M.E., Chapparo B.S., Data collection in the Palm of your hand: a case study. *International Journal of Human-Computer Interaction*. 13(2); 231-243. 2001.

Main D.S., Quintela J., Araya-Guerra R., Holcomb S., Pace W.D., Exploring patient reactions to pen-tablet computers: a report from CaReNet. *Ann. Fam. Med.* 2; 421-424. 2004.

[4] Greene P.D.,Handheld computers as tools for writing and managing field data. *Field Methods*. 13(2); 181-197. 2001.

[5] Pace W.D., Staton E.W., Electronic data collection options for practice-based research networks. *Ann. Fam. Med.* 3 (1); 21-29. 2005.

Wells T., Bailey J., Link M.W., Filling the void: gaining a better understanding of tablet-based surveys. *Survey Practice.* 6 (1). 2013. [http://surveypractice.org/index.php/SurveyPractice/article/view/25/html#r3]

[6] Peppard R., Belangia M., Pawlowicz R., Fattahipour M., Migues B., Attanayake P. Electronic capture of environmental data using a tablet PC mobile application for direct export to a database. WM'04 Conference, Bechtel, Technical Paper. 2004.

Beddall-Hill N.L., Jabbar A., Al Shehri S., Social mobile devices as tools for qualitative research in education: iPhones and iPads in ethnography, interviewing, and design-based research. *Journal of the Research Center of Educational Technology.* 7(1); 67-89, 2011.

Gutowsky L. F. G, Gobin J., Burnett N. J., Chapman J. M., Stoot L. J.,Blissa S., Smartphones and digital tablets: emerging tools for fisheries professionals. *Fisheries*, 38, 10, 455-461. 2013.

[7] ADB (Asian Development Bank), Using tablet computers for household surveys in Asia. ADB, Manila, the Philippines. 2013. [http://www.adb.org/features/going-geo-digital-household-surveys]

[8] Wei J., *Android Database Programming*. Packt Publishing, 2012.

[9] Wilson J., *Creating dynamic UI with Android Fragments*. Packt Publishing, 2013.

[10] Gamma E., Helm R., Johnson R., Vlissides J., *Design patterns: elements of reusable object-oriented software*. Addison-Wesley. 1994.

[11] Darwin I.F,. *Android cookbook - problems and solutions for Android developers*. O'Reilly, 2012.

[12] Linoff G.S., *Data analysis using SQL and Excel.* Wiley. 2007.

Industrial
Electronics
and
Engineering

Editor
Garry Lee