

Using CSPro (Census and Survey Processing System) Experience from Large Scale Surveys in India

Prabhu Ponnusamy

Abstract- This paper will discuss how CSPro (Census and Survey Processing System) coped with data processing system in complex large scale survey. A recent large scale surveys used the CSPro package for data entry fact sheet and tabulation generation for the survey. The paper will discuss the outcomes of using CSPro and data processing system methods in such a large scale survey. It is suggested that the use of CSPro has achieved a better data quality than other data processing packages would have. The use of CSPro has a number of distinguished advantages, such as improvements in data quality and turnaround times. It will critically review how the quantitative method worked in this specific situation before placing the discussion in its wider data processing system methods and research environment context.

Index Terms- CSPro (Census and Survey Processing System), SURVEY EXECUTING SYSTEM (CASES), DATA COLLECTION, LARGE SCALE DATA PROCESSING

I. INTRODUCTION

The purpose of this paper is to outline the experiences of the large scale surveys using CSPro (Census and Survey Processing System) package for data processing in India. The survey team used the CSPro package for a data processing during the conduct of the large scale surveys in recent times. This paper provides an overview of each phase of the data processing for survey from preparation to performance. It is hoped the information shared in this paper provides key insight for those national and state level survey agencies, researchers and statistical offices that may be considering using CSPro as an alternative method for data processing for surveys.

Main Phases of Data Processing

a. Preparation

- Software and hardware selection
- Develop data process cycle
- Training in use of survey software

b. Processes

- Data processing in large scale surveys
- Software and Hardware Selection

Software Selection

An initial concern was the size and complexity of the large scale questionnaire and software could cope with a questionnaire of that magnitude. The survey data collection instrument contained 5 modules namely Household, Ever married women, unmarried women, Village and Health facilities were used. The ability to

include smooth data process with national level to field agency and skip instructions, define values and incorporate rules were important criteria. The most important requirement however, was to obtain easy-to-use software for use by the questionnaire designers and ultimately the field agency data mangers and data entry operators.

After much research it as decide to develop in CSPro (Census and Survey Processing System) package for data processing and it was decide to develop in house IT - in-Charge with four member team in the project. The software has been used previously in much large scale globally using. Although it is the first in recent project Fact sheet and tabulation generation software also developed.

Why CSPro ?

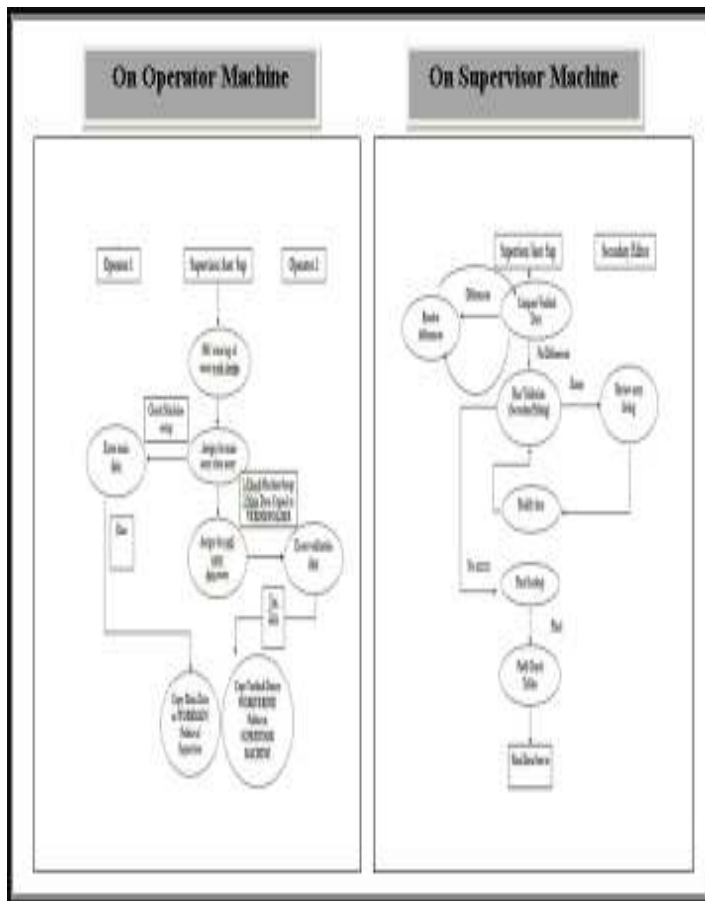
CSPro (Census and Survey Processing System) is a package for entry, editing, tabulation and dissemination of census and survey data. CSPro was developed jointly by the U.S. Census Bureau, Macro International and Serpro S.A., with major funding from the U.S. Agency for International Development. This software can be downloaded from <http://www.census.gov/ipc/www/cspro/> and is available free of charge. CSPro includes a data entry application, a batch editing application and a tabulation application. Within the data entry application there are facilities for defining the structure of your data (the data dictionary), and creating data entry forms. During the data entry process there has the facility to verify data by retyping values and comparing them to the values entered previously. It also has the facility to compare two data files created during double data entry, and produce a report on the discrepancies found. With the tabulation application you can produce cross-tabulations and frequency tables which are useful for exploratory data analysis and error checking. When you need to transfer to a statistics package to analyse your data further, CSPro provides a useful Export feature that transfers the data to formats readable by Excel and a variety of statistical packages. If requested this feature will generate syntax files for STATA, SPSS and SAS that contain the instructions for reading the data and for labelling the variables. The above mentioned distinguished advantages, such as improvements in data quality and turnaround times it has used for recent large scale surveys in India.

Hardware Selection

Machine should have at least P4 processor with 256 RAM and windows XP with higher version operating system. Better to format Hard disk before start of project. Install latest virus protection on the machine.

Develop Data Process Cycle

Total process in survey system can be divided in two parts Data Entry Machine and Supervisor machine the following is the process diagram will give clear picture.



The above listed instruction manuals is helpful guide field agency data processing in charge staff to smooth data process.

Training for Field Agency Data Processing In charge

A five-day training session was conducted for all field agency and monitoring agency data processing in charge staff. Roughly 50 persons were trained in understanding the purpose of the survey and how to effectively manage the data processing. The training also included questionnaire manage, survey data process cycle and handling error message.

Survey Data Process

Main activities of survey data process is listed below:

- **Receiving a Questionnaire from Field**
- **Office Editing**
- **Data Entry**
- **Verification**
- **Comparison**
- **Validation**
- **Coverage Report**

- **Field Check Table Generation**
- **Sending Data**
- **Fact Book/sheet Generation**

Once questionnaire of a PSU arrives in the central office, the PSU must be kept intact. One office editor will be responsible for editing and coding of questionnaire of a complete PSU, one data entry operator will enter all the questionnaires for initial keying of the PSU, a second keyer will enter the questionnaires for verification, and so on. Each PSU's data will be saved in a separate data file for that PSU, rather than into one large data file for the entire district. This is to protect against loss of data due to hardware or software failure.

After the data entry operator completes the verification, a report with discrepancies in both data file will be generated. The can be done after comparing both the files and correction must be made in verified data, which will be used for further process cycle.

The data entry application has been designed in such a way that it will be able to identify any inconsistency at the time of entering the data. However, there are some inconsistencies that require the attention of subject matter specialists or senior staff to resolve them. The result of executing this option is a report listing these types of inconsistencies. The report should be printed out and passed along with the package of PSU questionnaires to an editor. The second level editor will analyze the messages and decides whether data need to be changed or not. The second level editor will use the "Error Manual" to identify and properly resolve the inconsistencies. If the editor decides that no more changes are required or if after running the secondary editing no more messages are displayed. The data will kept in final folder, as soon as all the processes on a PSU data get over. He should update the log after uploading the data on file transfer protocol (FTP) site.

Coverage Report generated regularly, It gives two levels information on coverage, one at PSU level and other is at Questionnaire level. However, it also aggregates district figures at Questionnaire level. Field check tabulations generated as and when it requires by supervisor level staff in the field, monitoring agencies and Nodal agency Research Officer. These tables allow a review of the progress of fieldwork and help to identify problems with interviewers and teams. Corrective actions can then be taken.

Fact Book/sheet Generation will be created after completion the PSU of a Merging (Concat) 50 PSU's as district data on the final data, which will be kept in Fact book folder. For this multiplier and programs will be included in data processing software.

Conclusion

This was a first time in house developing software experience for testing the effectiveness of using CPro for data processing. The complexity of the recent Survey made for a challenging exercise, which in-house staff and field workers faced with innovation and determination. Upon reflection, it was quite evident that strict

survey controls were needed for the data process operations, such as weekly reporting to office staff, constant follow-up and increased observation of the recent survey data processing. However, the following outlines some of the benefits that can be assumed by using the CSPro package. Skip instructions, ranges and validation are built into the program in the data entry software provides greater security for data, improved data quality and large amount of survey data in limited time of delivery.

The recent survey optimistic in using CSPro as an efficient method of data process in large scale survey. As experience with the technology grows, its benefit hopefully can be exploited over a wide range of projects.

REFERENCES

- [1] Brewer, D.J. & Goldhaber, D.D. (2000). Improving longitudinal data on student achievement: Some lessons from recent research using NELS:88. In, D.W. Grissmer & J.M. Ross (Eds.), Analytic issues in the assessment of student achievement. Washington, DC: U.S. Department of Education.
- [2] Groves, R. M. (1987). Research on survey data quality. *Public Opinion Quarterly*, 51, S156-172.
- [3] Groves, R. M. (1989). *Survey errors and survey costs*. New York: John Wiley & Sons.
- [4] U.S. Census Bureau and Macro International, web <http://www.census.gov/ipc/www/cspro/doc.html>.
- [5] Statistical Services Centre, University of Reading web: <http://www.reading.ac.uk/ssc>
- [6] Quantitative research methods in educational planning
Web: <http://www.sacmeq.org> and <http://www.unesco.org/iiep>.

First Author – Prabhu Ponnusamy, B.Sc, M.A, M.Phil, Consultant State Data Manager, UNICEF Chhatisgarh, India and prabhustat@gmail.com.