

SECRETARIAT OF THE PACIFIC COMMUNITY

REGIONAL MEETING OF HEADS OF PLANNING AND HEADS OF STATISTICS (HOPS)  
(Noumea, New Caledonia, 22–26 July 2013)

**Session 5: HIES Field Based Data Capturing and Monitoring**  
(Document presented by the Secretariat of the Pacific Community)

**EXECUTIVE SUMMARY**

1. This paper describes a mixed data processing system called CAFE (*Computer Assisted Field Editing*) that aids in the monitoring and editing of the HIES with the overall goal of receiving the best possible data from the field by providing new tools to assess the performance of the enumerator and completeness of the HIES questionnaire.

**Key Outcomes**

2. Field monitoring – New tools and indicators were introduced to manage field collections and ensure that the study obtains 100 percent coverage of selected households within optimal time frames.
3. Data consistency and validity checks – Using programming logic to check the validity of data items and if those data items are consistent with other items in the survey. As well as developing the interface to easily update corrected data items.
4. Auto-coding and updateable reference databases – The PAC-COICOP database was integrated in the data entry application by implementing different methods of database indexing to accurately assign expenditure codes and implementing a system to update the PAC-COICOP database on country specific commodities.
5. Data management – Reducing the chances of data loss by limiting the management and transmission of data files from the user by automating the creation of data files, compression, restoration and transmission of those files.
6. Recode data into user-ready databases – The HIES database was recoded into different records. The benefit of having separate databases allows the end user to quickly analyse data specific to their needs without the hassle of searching over 200 data items.

**Recommendations for consideration by HOPS**

7. The new HIES CAFE has shown to be an efficient tool in capturing the best possible data from the field without compromising timeframes. It combines the mobility and flexibility of the paper questionnaire and data consistency checks of a CAFE system into a hybrid collection methodology that can be implemented by almost all NSOs willing to undertake a new and standardized approach in the collection of HIES data.
8. This paper recommends that *Pacific Island Countries and Territories* (PICTs) adopt the HIES CAFE for future HIES field work regardless of data entry location (in-the field, at headquarters) as evidence derived from the proposed new common Pacific HIES methodology applied thus far shows substantial improvements in data quality.

## INTRODUCTION

1. Objective 5 of the Ten Year Pacific Statistical Strategy (TYPSS) provided the framework for us to develop and introduce new statistical tools and systems. By designing and trialling new data collection methodologies and associated systems we are able to efficiently collect accurate and complete data from the field.
2. The trial of this new system was first implemented in Vanuatu during the collection of the 2012 Vanuatu Hybrid Survey. The unique collection methodology implemented during this survey relied on a year-round collection of data grouped into 24 one-week data collection rounds. During the one-week field collection the enumerator would visit a household multiple times. Each visit was used to collect information from six different modules. When a module was completed, the paper questionnaire was brought back to the supervisor to be keyed into a computer assisted field editing (CAFE) system. The CAFE system would then identify errors in the questionnaire and notify the supervisor of those errors.
3. The CAFE system was modified, improved and adopted in Household Income and Expenditure Surveys (HIES) in Nauru, Solomon Islands and just recently in the Federated States of Micronesia (FSM). This was possible because the field collection methodology and questionnaire design was similar to the Vanuatu Hybrid Survey but instead on 24 one-week collection windows the HIES was grouped into 16 two-week data collection windows.
4. This paper aims to highlight the achievements in terms of innovations and improvements. This paper will also provide the rationale behind the implementation of tools and the challenges associated in the development and implementation of the HIES CAFE.

## KEY OUTCOMES

5. *HIES CAFE Interface* – The management of HIES data collection consumes many resources. Taking this into consideration the role Data Quality and Control (DQC) Officer was elevated to encompass the roles of the supervisor, entry operator, data coder and control clerk. A system was created that allows the DQC officer to easily navigate and execute the tools provided in the HIES CAFE to enter data, check the progress of the survey, produce error reports, code data items and transmit data.
6. *Field monitoring* – Four utilities have been implemented to allow the DQC officer to monitor field work. The first being the management of the sample to prevent duplicate case and identify missing or unaccounted households. The second utility records the number of errors identified and aggregates that information either by enumerator, supervisor or enumeration area. The third utility records the number of modifications made to the case to monitor data entry times and the fourth utility records the number of days between the interview and data entry.
7. *Data entry* – The data entry screen in HIES CAFE mimics the layout of the questionnaire to a certain extent. At the same time allowing individual data items some room for modification with relative ease when implementing the HIES in different countries. When converting the questionnaire entry screen to a different language an Excel macro was developed to allow a PICT to make those changes without going into CSPro entry design. Thus reducing the amount of time to implement the HIES CAFE into a new PICT.
8. *Data consistency and validity checks* – One of the main features of the HIES CAFE is the ability to check the quality keyed data. This was made possible by implementing four different error/consistency checks.
  - a. The first is that the data entry program must be able to check if the value entered for each data item is within the scope of expected values. For example, the possible values for the SEX item is "1" to represent a male and "2" to represent a female. If the DQC officer accidentally puts a "3", the SEX item contains a value that is outside the scope and the data entry

- program should display an error message informing the supervisor that the value entered for the SEX item is invalid.
- b. The second type of error that the data entry program should look for is consistency of data items across multiple variables. Using the example of the SEX item, if the DQC officer enters a "1" for a male member of the household a few questions later the DQC officer enters a valid value for number of births which is a question that is intended for a female respondent. The data entry program should display an error message that the value entered in fertility, references a male respondent.
  - c. The third error type applies to value amounts. The third error is not necessarily an error but a warning advising the DQC officer that a value entered for a particular product or service may be a bit too high. This calculation is dependent on the item, the amount paid and the frequency of payments.
  - d. The fourth error check is built from the expenditures and income recorded from the modules and diaries. This process sums all values then properly identifies and categorizes the groupings into disbursements and receipts. The result is a table of 14 different types of disbursements and 20 different types of receipts along with the percentage of that indicator over the total amount of disbursements or receipts. These percentages are then cross checked with established thresholds allowing the DQC officer to quickly identify doubtful values in the questionnaire.
9. Another feature that was built into the data validity and consistency checks is the ability to correct the errors encountered without the need to navigate over 90 pages of information. When a data error is encountered the HIES CAFE interface allows the DQC officer to navigate to the exact question to make the correction as well as keeping track of those changes. The result is a systematic process of correcting mistakes thus producing quality data quickly and efficiently.
10. *Auto-coding and updateable reference databases* – to accurately assign PAC-COICOP codes to expenditure items, a lookup system was developed and integrated into the data entry program. The program provides two ways for the DQC officer to quickly and accurately assign a PAC-COICOP:
- a. The first is a key-word search feature that searches a PAC-COICOP database for commodities that contain the specified a search term.
  - b. The second search feature takes advantage of the PAC-COICOP hierarchy allowing the DQC officer to filter the correct code depending on its purpose. If a PAC-COICOP code is not found the DQC officer can add a country specific PAC-COICOP using the update feature integrated in the HIES CAFE interface.
11. *Data management* – To reduce the chances of data loss, the HIES CAFE controls the management and transmission of data files in three different processes.
- a. The first process creates data files based on established naming conventions making it easier to identify the source and version of the data file.
  - b. The second process creates backup of the by compressing all data and information files and allowing the DQC officer to save the information onto a removable storage or local hard drive.
  - c. The third process allows the DQC officer to send a compressed transmission file via file transfer protocol (FTP) to a central server or gives the option for the DQC officer to save the transmission file to a removable storage device. The program can be amended to invoke the default email client and send the transmission file via email to the NSO.
12. *Recode data into user ready databases* – A major improvement implemented into the HIES CAFE is the ability to recode the HIES data into user ready databases. That can be used with relative

ease to produce NMDI indicators and subject specific indicators with having to search a multitude of data items.

- a. Household's Basic Data - contains information about the characteristics of the dwelling.
- b. Household's Housing, Income, and Expenditure Summary Data – contains information on the expenditure and income of the household. This record also includes wealth index deciles and poverty indicators.
- c. Household Member's Data – contains basic demographic, education, health, ICT and labour force participation information.
- d. Household's Detailed Annual Expenditure Data – contain information on the household expenditures on the commodity or service level and can be used for CPI weights.
- e. Household's Ownership of Major Appliances and Vehicles
- f. Household's Expenditure Data from the Diary – contains information on the household's expenditure patterns indicated from the diaries.
- g. Household's Annualized Expenditure Data from the Modules – contains information on the household's expenditure patterns indicated from the modules.

### **MAJOR CHALLENGES**

13. Adding country specific questions – Because the interface, recode and error program was designed from a standard questionnaire it will be difficult to add country specific questions into the middle of certain sections or modules. There will be a need to train dedicated NSO staff in the modification of the HIES CAFE to tailor the system to the NSO needs. Having an agreement across the region on the standard set of questions and proper ways to administer the questions will help alleviate problems associated in the modification of the HIES CAFE.
14. Possible dependency on the system for quality control – The HIES CAFE was designed to help correct current and potential errors that can arise from field collections by systematically providing a method of error and consistency identification. This may cause a dependency of the DQC officer, supervisor and enumerators on the system to identify such errors. The fact is that although there are tools in place to obtain the best data as possible the first line in excellent data quality is well trained and proactive field collection team.

### **RECOMMENDATIONS TO HOPS**

15. The new HIES CAFE has shown to be an efficient tool in capturing the best possible data from the field without compromising timeframes. It combines the mobility and flexibility of the paper questionnaire and data consistent checks of a CAFE system into a hybrid collection methodology that can be implemented by almost all NSOs willing to undertake a new and standardized approach in the collection of HIES data. It contains innovations that has not been used in other HIES across the globe and we are confident that this tool will be invaluable in future HIES collection activities.
  - i. This paper recommends that PICTs adopt the HIES CAFE for future HIES field work regardless of data entry location (in-the field, at headquarters) as evidence derived from the proposed new common Pacific HIES methodology applied thus far shows substantial improvements in data quality.