<u>SECRETARIAT OF THE PACIFIC COMMUNITY</u>

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

Session 5: New and innovative statistical tools and systems have been introduced - Implementation of Pacific Living Condition (Hybrid) survey prototype in Vanuatu
(Document presented by the Secretariat of the Pacific Community)

EXECUTIVE SUMMARY

1. Development of new household survey prototype, supported by the World Bank Trust Fund for Statistical capacity building, that:
   a. could be used across the Pacific and thus provide access to comparable development statistics across key sectors as requested by Pacific Leaders in the original Pacific Plan;
   b. combines core elements from major Household surveys, thus providing countries with access to key national and international development indicators, required for monitoring and reporting on policy performance and development progress; and
   c. is able to capture key statistical development indicators in one survey, reduces data collection costs, facilitates more frequent collections thus providing feedback on policy performance and development progress on a more regular basis than every 5 years.

Key achievements

2. *Questionnaire design* includes comprehensive set of regionally/internationally comparable indicators (82 *National Minimum Development Indicators* (NMDI) indicators) relevant to national policymakers and international reporting requirements; it also includes a solid measure of consumption that allows analysts to link each of the sectoral indicators to household income and poverty.
3. *Data collection* amongst 1,200 households spread over 12 months to address *seasonality,* with *data management* integrated into fieldwork to permit computer-based quality controls during data capture (*Computer Assisted Field Edit*).
4. *Data analysis* currently underway, which will deliver 5 fact sheets by end of August (household cash expenditure; household subsistence activities; poverty; perception of well-being; cultural practice indicators), followed by labour force, health, and nutrition profiles by the end of the year.

Main Challenges faced

5. Good field supervision remains challenging, particularly when required to recruit province-based staff without much prior survey experience and computer skills (*see Session 1*).
6. Initial computer assisted field edits hampered by lack of regular access to power, which was resolved by purchase of small generators, only to face another problem with transport (as they cannot be loaded on planes).
7. Data transmission to the central office in Vila remained problematic throughout, as most provincial supervisors lack access to internet.

Recommendations

8. To review management procedures regarding management/supervision of field supervisors.
9. The application of Computer assisted field edits has the potential to tangibly improve survey data quality, but needs to be backed up with adequate logistics and IT support.

KEY ACHIEVEMENTS

Questionnaire design

1. The design and subsequent testing in Vanuatu of the Pacific Living Conditions survey, also known as the Hybrid survey, which combined key thematic ingredients from existing Household surveys, notably HIES and DHS, was a logical step forward in developing a (light) modular survey instrument that could become the cornerstone of a regionally-coordinated regular survey program,
   a. providing a core set of comparable development statistics and minimum indicator datasets for all countries as envisaged by Pacific Leaders,
   b. capturing the required input data to generate standard HIES and DHS outputs, and given its modular design,
   c. provide opportunities for countries to customize their surveys by including customized modules to address specific national (and regional) information demands.

2. Combining the core elements of the HIES and DHS, in terms of providing countries with key development (e.g. MDG) indicators, this survey will be less expensive than running two fully-fledged HIES and DHS to merely collect statistics than cannot be generated (yet) from administrative databases and management information systems. And running it in shorter time intervals (every 2-3 years) would provide countries with better quality data for progress monitoring.

3. This modular structure, also lends itself to lower survey costs/increase cost-efficiency, by providing a platform for other government agencies to buy space on this survey platform every 2-3 (or 4-6) years, without having to pay extraordinary costs to mount a household survey on its own. Paying greater attention to, what used to be called *Omnibus*-surveys, will lower both survey costs and survey fatigue (amongst the general public and NSO staff).

4. Questionnaire design includes comprehensive set of regionally/internationally comparable indicators (82 NMDI indicators) relevant to national policymakers and international reporting requirements; it also includes a solid measure of consumption that allows analysts to link each of the sectoral indicators to household income and poverty.
   a. *HIES component* (income and expenditure): 2 modules of expenditures based on recall period, and a 7 days diary. The income part is covered in a specific income module (recall).
   b. *Health/Demographic component*: in addition to the expenditure on health covered in the HIES component, the hybrid survey collect DHS-type information on:
      a. Specific health indicator on children less than 5 (weight, height, vaccination)
      b. Health status of all the family members (chronic illness, sickness, access to health facilities)
      c. Pre natal and post natal cares for pregnancy
      d. Birth history, infant mortality, fertility
      e. Family planning, contraceptive use
   c. *Labour force:* the labour force module records all main activities household members aged more than 10 years were involved in during the last 12 months. All individuals ten years and older are eligible for this section and need to complete this form, even if they were not involved in any paid work/business. Student, house workers, subsistence worker are recorded. We allow for each household member aged 10+ to recall up to a maximum of 4 activities they were involved in during the past 12 months.
   d. Three additional modules were included to explore the viability of collecting other sectoral information (cultural practice, nutrition), and to undertake a test of recall method versus diary capture.
      i. *Recall on food expenditure*: inclusion of module to test viability of recall-method on cash expenditure on food over the past week, covering 53 main food items usually purchased by Vanuatu households

ii. A one-week *Food Diary* was included in this survey, to establish dietary patterns of ni-Vanuatu households, particularly of school-age children and people in the labour force, to ascertain the compatibility of "fuel' intake (food-groups) with a day's learning and economic activity.

iii. *Cultural Practice* module. With culture representing a strong foundation of sustainable development in many Pacific island countries, yet little factual information available on various manifestations, we developed a module of "cultural practice" for inclusion in this survey, to be able to capture three basic manifestations of cultural practice, as reflected in SPC National Minimum Development Indicator database:

- indigenous language skills and use,
- traditional production skills, and
- people's access to and use of customary land, forest and marine resources.

Sampling

5. The hybrid sample size comprises of 1,200 households, which is smaller than other usual HIES survey sample size in Vanuatu (3000+ households). This sample size only provides breakdown at urban/rural level, and even if EAs from each province are selected in the sample, and especially in the rural area, this sample size is big enough to provide provincial outputs.

6. The sample was selected in two stages:
   - In the first stage, 100 Census Enumeration Areas (EAs) will be selected with probability proportional to size from the 2009 Vanuatu population census; and
   - In the second stage, 12 households will be selected with equal probability within each of the chosen EAs.

7. Stratification was done in each province according to the area:
   a. Urban area:
      - Urban (Port Vila and Luganville)
      - Peri –urban (villages around Port Vila and Luganville with easy access to town)
   b. Rural area:
      - Provincial centre (e.g. Lenakel in Tafea, Lakatoro in Malampa)
      - Rural (North Efate in Shefa province)
      - Far/Remote Rural (other islands, or not easy to access: e.g. North West Santo)

   Budgetary consideration means exclusion of some rural EAs from sample, as they were too remote and expensive to reach (e.g. Torres islands).

Fieldwork

8. Field operations where scheduled over a twelve month period, starting in February 2012 and running to the end of January 2013. Thus allowed covering for seasonality effects, as they pertain to expenditure patterns, which are associated with work opportunities and income (including, of course, seasonal agricultural work).

9. This 12 month period of field operations was divided into 24 survey rounds; a survey round covers a 2-week period dedicated to the interview of a sub sample of the total sample. During this period, each survey team covered 12 households in the same EA, with 6 households allocated to each enumerator. After completion of each round, each team moves on to another EA in the same province.

ONE ROUND OF COLLECTION

	WEEK 1 (W1)							WEEK 2 (W2)						
	W1 Day1	W1 Day2	W1 Day3	W1 Day4	W1 Day5	W1 Day6	W1 Day7	W2 Day1	W2 Day2	W2 Day3	W2 Day4	W2 Day5	W2 Day6	W2 Day7

HH1, HH2 and HH3 = 7 days of diary
HH4, HH5 and HH6 = 7 days of diary

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
HH1 HH2 HH3	HH4 HH5 HH6	HH1 HH2 HH3	HH4 HH5 HH6	HH1 HH2 HH3	HH4 HH5 HH6	Rest	HH1 HH2 HH3	HH4 HH5 HH6	- Final checks - Prepare for next round			- Rest	
Visit1		Visit2		Visit3			Visit4						

Visit1	Visit2	Visit3	Visit4
- Presentation - Module 1 - Dairy presentation - Fill the first dairy day	- Diary check - Module 1 feedback - Module 2 & 3	- Diary check - Module 2 & 3 feedback - Module 4	- Last diary checks - Module 4 feedback - Module 5 & 6 - Final global checks - Final thanks

10. Each team follow different schedule according to their province:

From February 2012 to January 2013
24 rounds = 1 year

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20	R21	R22	R23	R24
Torba 1																								
Sanma1																								
Penama1																								
Malampa1																								
Shefa 1																								
Shefa 2																								
Tafea 1																								

☐ Field operation on for that team during that round

11. Apart from being able to address the impact of seasonality, this survey design also allows field teams to review their work after each round and thus contribute to improving the quality of their work output round after round.

12. Field staff were organised into 7 provincial teams, with each team comprising of three staff: one supervisor in charge of data coverage/quality and for the survey to proceed along schedule, and two enumerators, each in charge of 6 households, to undertake the interviews and regular diary checks. In total, 30 staff were trained, 21 deployed in the field and 9 reserves.

13. Working with a small number teams had the following advantages:

- fewer staff to recruit and train (and allow for better quality of training)
- Better quality of the field staff monitoring

Data management/Data entry

14. In each team, the team leader (supervisor) was in charge of data entry. Each team was provided with a laptop, and data entry of each round was done during the field operation of that particular round. The data entry system was designed in such a way as to include feedback to the supervisors through an error listing. This report lists all inconsistencies recorded on each form, and all missing values. From this report, the supervisor is able to assess the quality of the enumerator's work, and monitor it closely.

15. This system has two obvious and tangible advantages:
    ▪ it provides immediate feedback on the quality of the job done by each enumerator;
    ▪ edited dataset are available straight after completion of the field operation.

Data dissemination

16. Survey outputs will be released in the form of thematic 4-page factsheets, providing a summary of main findings, trends, and urban-rural differentials. For the moment, five thematic factsheets are planned to be released by the end of August addressing:
    a. household cash expenditure
    b. household subsistence activities
    c. poverty
    d. perception of well-being, and
    e. cultural practice indicators.

This will be followed by similar factsheets on labor force, health, and nutrition by the end of the year.

MAIN CHALLENGES

Field-based data entry

10. Notwithstanding careful planning and the development of computer-based quality controls during data capture (*Computer Assisted Field Edit*), some serious obstacles became quickly apparent:
    a. Computer literacy of some supervisors was very limited;
    b. in many EAs access to power was either limited (to a few hours a day) or did not exist, and once small generators were purchased, realization dawned that they could not be transported as cargo in light aircraft; and
    c. accommodation limitations in the village with no proper space for the supervisor to work or sit while doing data entry and verification checks.

11. To be efficient, data entry was meant to happen at the end of each day, after completion of interviews, to allow for immediate consistency/coverage checks and edits, and be able to inform the enumerator what corrections needed to be made during his next visit. This was the plan, and instructions provided during training.

12. Reality in the field, however, was different as illustrated in the graph below, with the average time lag between interview and data entry varying between 1.2 days for the Malampa team, and more than 5 weeks for the Penama team!

**Average period of time (in day) between interview and data entry**
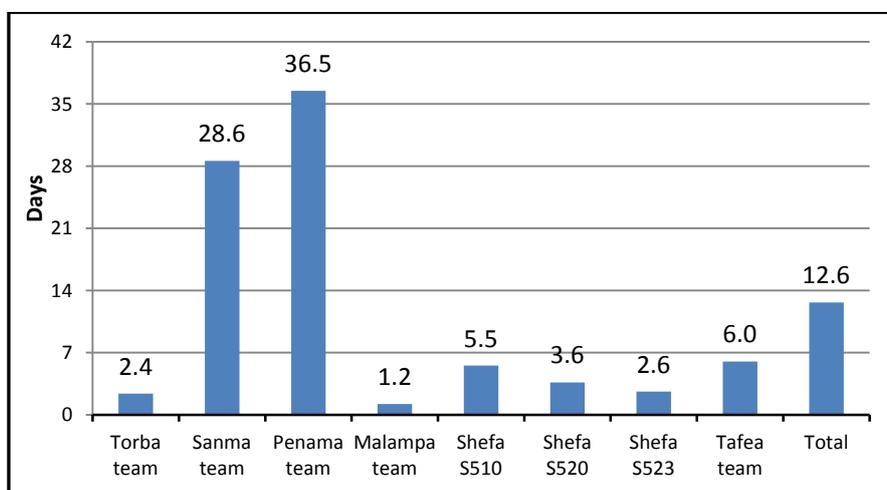
Note: Both Shefa 520 and Shefa 523 make up the Shefa-2 team.

13. Good results were achieved by the Malampa, Torba and Shefa-2 teams, whereas Shefa-1 and Tafea teams averaged a time lag of between 5-6 days after the interviews were completed; this is too long as with the team staying less than 10 days, this period of time does not allow proper field edit, check and correction on the spot.

Obstacles to field-based data entry

14. Thanks to the power generator, data entry in rural areas was efficient in Torba, Malampa, Shefa2 and Tafea. We can clearly see that access to power was not the main issue for data entry in Sanma and Penama, as both supervisors took more time to achieve data entry in the EA with easy access to power.

15. In most of the rural EAs, supervisors had no access to a reliable power source in order to keep their laptops charged. Once this was brought to the attention of the Port Vila based survey manager after a few survey rounds, a small generator was purchased for the rural survey teams. While this improved data entry markedly in rural Shefa, with no major differences emerging in Tafea, Torba and Malampa, the graph below shows quite clearly that access to power was not the main impediment for slow data entry in Sanma and Penama provinces, with both supervisors taking more time to complete data entry in the EA with easy access to power.
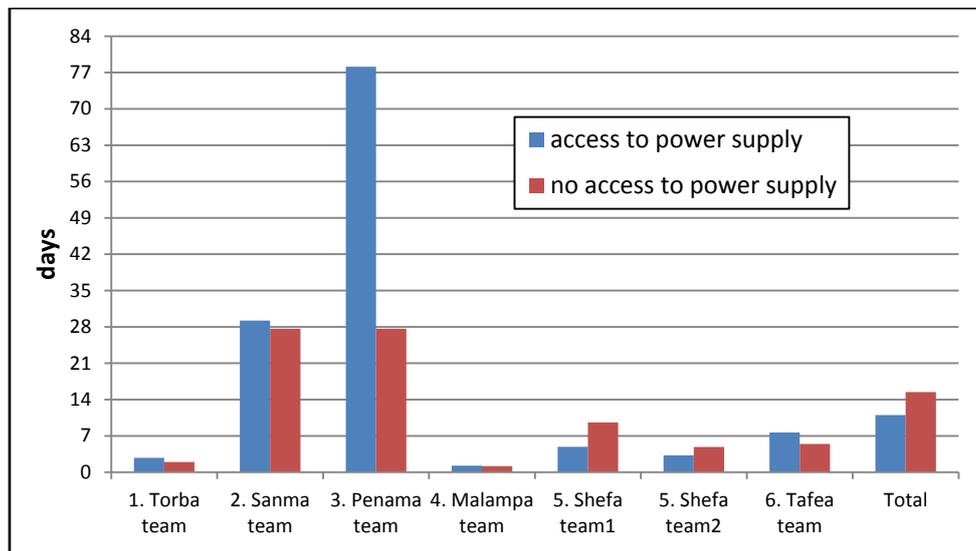
16. A combination of limited computer skills and perhaps 'attitude' by the Sanma and Penama supervisors, combined with less than ideal field-management/supervision, and not enough IT support from NSO might provide a more realistic assessment of the very slow and inefficient field-based data entry in these two provinces.

Data transmission

17. The overall methodology is efficient as long as the datasets, after field-based data entry are centralised in the NSO, processed and a feed-back report is provided to the team – representing an ideal way of performance monitoring the field operations. In many cases, however, sending data to Port Vila proved to be a real challenge for the provincial teams, especially from Torba, Tafea, Penama and Malampa, the two main obstacles being:
    ▪ limited access to the internet to forward the data through email, or
    ▪ no access to an airport to send data through a flash drive backup.

Data quality

18. Data quality assessment is based on 3 main indicators:
    - Number of days between interview and entry (as low as possible)
    - Number of items collected in the diaries (if there is de downward trend after a few rounds the enumerators might miss some diary check and under report the record of items)
    - The number of errors reported by the system

19. The whole purpose of field-based data entry was to improve data quality, with teams encouraged to edit field data straight after data entry, to be able to spot and correct mistakes and complete missing values. On average, the system reported 4 errors per questionnaire and a time lag of 13 days between data entry and interview, with again substantial variations emerging between the provincial teams (see graph below).
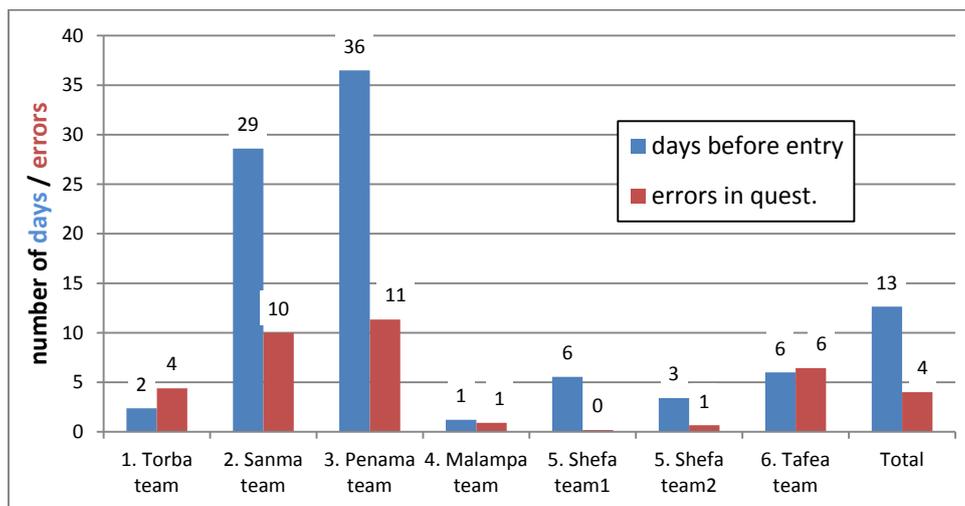
LESSONS LEARNED/RECOMMENDATIONS

20. With field-based data entry providing the biggest challenge for all field teams, and reflecting remarkable contrasts in performance and quality between provincial teams, greater attention needs to be paid to recruitment and training, with field supervisors/data entry operators expected to have an acceptable standard of computer literacy. The lack thereof would transpire during training and an end-of-training data entry proficiency assessment.

21. The application of Computer assisted field edits has the potential to tangibly improve survey data quality as has been shown in this survey, but clearly needs to be back up with adequate logistics and IT support, and perhaps a more intense, and assessment-based data-entry training.

22. As has been reported with censuses and other household surveys earlier, general field supervision remains a challenge. More regular field visits by NSO staff might have picked up performance inefficiencies earlier, and changes could have been made, given a reserve list of 9 staff.

23. Given the overall poor performance of the Sanma and Penama teams, the NSO might wish to reflect on this and on how to improve statistical operations in these provinces.