



# ***Spotlight with NRI***

## **Development Issues, Policies, and Trends**

October 2010

Vol. 4 No. 5

### **Attributes of reliable data collection: Some hints for Papua New Guinea**

*Israel Sembajwe*

#### **Introduction**

In the second quarter of 2009, the National Statistical Office (NSO) was working on the final stage of reviewing, printing and disseminating the 2006 Papua New Guinea Demographic Household Survey (DHS) Report. It was an honour for those who participated in this final leg of the survey to make a small contribution to the review. However, as the review was being concluded, many questions came to the minds of those involved.

One of the questions was, “If the analysis report of the 1996 PNG DHS was released in 1997 (only one year after the survey), why did it take three years for the NSO to finalise the 2006 PNG DHS Report?”

These questions were informed and justified by the 2009–2010 Papua New Guinea Household Income and Expenditure Survey (HIES) process. The enormous logistical and financial constraints which the NSO faced in executing this survey and the implications of these challenges on the reliability of the expected data were worrisome, as outlined by Alvaro in his 2010 World Bank mission report.

There is a need to put the spotlight on the attributes of reliable data collection so that future mistakes can be avoided. This is particularly important at this juncture in time when the NSO is preparing to carry out all the activities related to the next national census, which has been now postponed to 2011.

**The National Research Institute—developing policies for national development,  
good governance, and more informed decision making**

## **The importance of data and information**

In his 2006 paper titled, *The Importance of Accurate, Reliable and Timely Data*, Eslake stated that, “In the absence of accurate, reliable and timely information, people and organisations will make bad decisions; they will be unable to help or persuade others to make better decisions; and no-one will be able to ascertain whether the decisions made by particular individuals or organisations were the best ones that could be made at the time.” Eslake emphasised that “accurate, reliable and timely information is vital to effective decision making in almost every aspect of human endeavor, whether it be undertaken by individuals, communities, organisations, businesses or governments.”

It is, therefore, very important that governments and government agencies which are tasked with the important role of data collection take ownership of the process with extreme seriousness. In doing so, they should know that careful planning is critically important to a successful data collection exercise. This is particularly so in the most important national data collection exercises, such as the census. It is not only important in countries with comparatively little statistical experience but also in those with well-developed systems of statistics. Failure to do so leads to grave decision-making and planning consequences. These issues are discussed in detail in the United Nations recommendations on population and housing censuses.

In particular, in a situation where a data collection exercise is badly handled, subsequent arguments revolve around the integrity of the data, rather than the facilitation of decision-making processes for which the data were intended to support. This is extremely bad business and it retards a country’s progress.

### **Attributes of reliable data collection**

In the case of the census, it should not be just the desire to fulfil a routine of carrying out a census every five or ten years, but the desire to fulfil a commitment attached to the importance of collecting such data by the government and government agencies. This commitment is reflected by the keeping of responsibility and showing accountability by all actors. It is also demonstrated by mutual respect, recognition of roles played, and timely facilitation.

These ingredients can be achieved by the provision of a comprehensive work plan covering all stages of the exercise from pre-enumeration to data

processing, analysis and dissemination; determination of realistic timelines; provision of an holistic comprehensive budget (which may experience some moderate adjustments as the process progresses); well-organised logistics and quality control mechanisms; engagement and training of adequate and suitable human resources; and engagement of all partners in the process.

Let us now reflect on some aspects of these ingredients as we briefly outline the attributes of reliable data collection, using the census as an example. The census is the most important, extensive, complicated and expensive statistical operation in any country. Handling it properly provides data that can be disaggregated to the smallest administrative unit, as well as creating a rich list of enumeration areas that can provide a good base for a sampling framework that facilitates the execution of national surveys.

The selected attributes are as follows:

### **General information**

Census preparation can be generally divided into six phases: (a) preparatory work; (b) enumeration; (c) data processing; (d) building of needed databases and dissemination; (e) evaluation of the results; and (f) analysis of the results. In addition, distinct sets of operations related to the systematic recording of census experience and the quality control and improvement program must accompany and support the main census operations. These phases are interrelated.

### **Preparatory stage**

The preparatory work for the census is long and involves the following range of activities:

- legal basis for a census;
- budget and cost control;
- census calendar;
- administrative organisation;
- communications activities, including consultations with users and census publicity;
- plans for the quality control and improvement program;
- cartographic work;
- small area identification;
- living quarters and household listing;
- tabulation program;
- questionnaire preparation;
- census tests;
- plan of enumeration;

- plans for dissemination;
- staff recruitment and training;
- avoiding gender biases and biases affecting data on minority populations.

### ***Calendar***

A timetable indicating the sequence and estimated duration of each of the component operations of the census is the foundation of a carefully planned census process and is called a *census calendar*. At the early stages of census planning, a provisional calendar of selected key dates should be prepared as an overall framework for the census, and this is revised and made more detailed as planning proceeds.

Care should be taken to ensure that no major statistical inquiries are occurring simultaneously and competing for resources.

### ***Administration***

National, subnational and local commissions and committees are frequently useful in the planning and preparations of a census and should be timely organised. Such bodies may be composed of representatives of governmental agencies and non-governmental users of the census data.

### ***Advocacy and publicity***

Advocacy and publicity of the census cover three distinct audiences: (a) major users of census data; (b) persons and institutions participating in the census operations; and (c) the general public. It is through advocacy and publicity that governmental and non-governmental organisations outside the census office may be called upon to provide various types of assistance to the census process.

### ***Quality control***

Quality control mechanisms are put in place to detect errors so that remedial actions can be taken as the census process continues. If no such mechanisms are provided, the generated census data may contain many errors which can gravely affect the final results making them questionable. In addition, the mechanisms should cover all phases of census (including planning, pre-enumeration, enumeration, document flow, coding, data capture, editing, tabulation and data dissemination).

### ***Cartography or mapping***

Up-to-date maps are needed to set up enumerator assignments, estimate travel time and costs, establish field offices, assign geographical codes, determine the best route of travel to and within enumeration areas, measure distances and enable the field staff to locate an enumerator or find a specific housing unit when a return trip is neces-

sary, among other things. Therefore, it is necessary to start cartographic work sufficiently early to ensure that an adequate supply of maps is available three to four months before the census is scheduled to begin. For population and housing censuses, the lead time for mapping should be at least two or three years, even when there is an ongoing mapping program.

### ***Listing***

A list of sets of households that are available at the start of the census is an aid in the control of the enumeration, particularly in the absence of adequate and updated maps. Such a list is also useful for estimating the number of enumerators and the number of schedules and other census materials needed in an area, for estimating the time required for the enumeration and for compiling provisional results of the census.

### ***Tabulation program***

It is essential that the tabulation program is outlined sufficiently early so that the procedures and costs involved may be investigated thoroughly before a final decision is reached. The testing of the questionnaire will help indicate whether gathering the material desired for tabulation is a reasonable possibility.

### ***Questionnaire***

The type of questionnaire, its format and the exact wording and arrangement of the questions merit the most careful consideration, as the handicaps of a poorly designed questionnaire cannot be overcome during or after enumeration. Among the many factors that should be taken into account in designing the questionnaire are the method of enumeration, the type of questionnaire, the data to be collected, the most suitable form and arrangement of the questionnaire and the processing techniques to be employed. Many decisions regarding processing are dependent on the final content, form and arrangement of the questionnaire.

Questionnaire design should be considered jointly with the planning of the tabulation program to ensure that information needed for the tabulations will be available. The final questionnaire must be drafted in time to allow for proper training of census officials, for adequate publicity to be generated on its content and for any delays in printing.

### ***Census tests and census pilot***

The testing of various aspects of a census plan prior to the enumeration is useful for determining the adequacy of the questionnaire, the field organisation, the training program, the processing plan and other important aspects of the census.

In the case of the questionnaire, the tests help to verify the suitability of intended census questions, including their formulation and the instructions provided, as well as the suitability of the questionnaire design. These tests are also used for estimating the time requirement in enumeration. It is practical to carry out questionnaire tests on a small scale in several purposely selected places. Because they are relatively inexpensive, repeated rounds of questionnaire tests may be carried out until a satisfactory questionnaire has been evolved.

On a bigger scale, a comprehensive test of all census procedures is a *pilot census*. The pilot census tests the adequacy of the entire census plan and of the census organisation. In order to best serve this purpose, it should be undertaken in conditions resembling the actual enumeration as closely as possible. For this reason, it is often taken exactly one year before the planned census so as to conform with the expected seasonal patterns of climate and activity.

#### ***Other important aspects***

A good comprehensive census plan should also take care of the following:

- plan of enumeration;

- plans for data processing;
- plans for analysis and dissemination;
- plans for staff recruitment and training; and
- plans for avoiding gender biases and biases affecting data on minority populations.

All these aspects, including all those discussed earlier, contribute to reliable data collection.

#### **Conclusion**

These aspects of attributes for reliable data collection suggest a need for reflection on whether adequate deliberation and planning have been conducted in preparation for the next census. If not, what do they need to do to achieve a census process that produces reliable data for the important exercises of national decision-making and planning? Pitfalls that have been revealed by the PNG DHS and PNG HIES should be avoided and best practices from past data collection exercises be identified and utilised.

#### **About the Author**

*Dr. Israel Sembajwe is Professorial Fellow and Head of the Social and Environmental Studies Division at the National Research*

**I**

**Contributions to Spotlight should be addressed to the Director, NRI:**

**Email: [nri@nri.org.pg](mailto:nri@nri.org.pg)**

**Fax: 326 0213**

**Post: P.O Box 5854, Boroko. NCD 111. Papua New Guinea**

**Further information on the National Research Institute can be obtained from the Institute's website—[www.nri.org.pg](http://www.nri.org.pg)**