

# **DATA PROCESSING OF THE 1999 POPULATION CENSUS IN VIET NAM**

Prepared for the ESCAP Expert Group Meeting on  
Effective Use of IT in Population Censuses  
Bangkok, 10-12 December 2007

## ***1. Census history***

The first real population census of an independent Viet Nam was conducted late in 1979. Given the resources and technical skills available at that time, the census provided surprisingly good benchmark data as a springboard for national level.

The first census which could really be considered as a modern census, introducing internationally recognized census concepts, design features and processing, was conducted in April 1989.

The third census was conducted in April 1999. As will be seen, many of the features of the 1989 census were incorporated into design and conduct of the 1999 census. In addition, however, the 1999 census added new questions and extended its scope in some areas to provide even more comprehensive data. The two censuses together will provide a rich source of data to analyze the current situation and key trends over past ten years.

The next census is planned to be conducted in 2009. At this moment, the preparatory activities are well underway.

## ***2. Data processing of the 1999 census***

### ***2.1 Selection of software***

Questionnaires used in the main 1999 census were designed in booklet format, where separate columns were provided across each page to record information for each person. While these booklets proved very useful in recording and checking information, they presented some problems in data processing and in selecting and designing software for data entry. In late 1998, the Central Data Processing Centre (CDPC) conducted test runs to develop its data processing system, using records from the pilot census. Three different applications were tested for data entry (adapting standard packages known as IMPS, ISSA and FOXPRO). From these tests it was decided that the version based on ISSA was the best-suited to the Viet Nam census.

## *2.2 Distribution of data processing facilities*

Data entry and on-line editing facilities were provided at 9 centres (CDPC, Hanoi, Nam Dinh, Da Nang, Khanh Hoa, Binh Thuan, Ho Chi Minh City, An Giang and Can Tho). Each centre established a computer network for census data processing with a server HP LH3 and 12 to 53 PCs. For the entire country, 240 PCs were used for processing and 10 servers (two in CDPC). A total of 450 data entry operators were engaged to work in two shifts. The network in each regional computer centres is connected to the CDPC network through a telephone dial-up system. Thus, as data entry was completed in each province, data files were transferred to the CDPC server.

## *2.3. The data entry, editing and tabulations operations*

The main strategy for data entry is to "photocopy" all information on the questionnaires to the computer, that is to change recorded information as little as possible. To do this keyers were not permitted to modify or correct recorded information. However, there were a number of checks provided to ensure the keyed data were valid, but most of these checks were in the form of warning messages to control for key stroke errors and column shifting. Data entry was completed for the 3 percent sample within about two months (working July and August 1999). All editing operations were undertaken at the CDPC. Tables were produced progressively from September 1999, using the CENTS module.

A number of consistency checks were carried out (the edits were specified in CONCOR) and records in error were edited on-line by special editors. As soon as data entry for a district was completed, a listing of inconsistencies was printed out for checking and, correction and updating of files.

## *2.4. System management and control*

A system of management and control was developed by Central Data Processing Center (using a VISUAL FOXPRO application) to help managers to monitor the processing. All stages in the cleaning of data are monitored under this system, from the receipt of the questionnaires, to data entry, verification, checking, listing of inconsistencies, data correction, combining EA data files into higher geographic levels, production of frequency tables, and data backup. The system also provides the mechanism to validate the geographic identification of keyed data, to avoid duplication or omission of EAs. For managers, the system generates different kinds

of reports, for example, to keep track of the status of each EA, to calculate the quantity and quality of work of the data entry operators and print salaries due, or to provide frequencies of imputed values to subject matter specialists to ensure the rules were properly applied.

### *2.5 Data Dissemination*

Given the very high interest in the results of the census, an important objective in processing the sample was to produce tables as soon after completion of fieldwork as possible. In general, the strategy adopted gave rather less weight to procedures for eliminating all possible sources of error than would be given for the full results, since it was assumed that users would be willing to accept small errors in return for the benefit of receiving a comprehensive range of tables within a year of the census. But nevertheless it should be appreciated that small difference might appear between the early sample figures and those will be published after the entire census has been processed.

Several specific steps were taken to speed up the processing of sample forms. Sample enumeration areas (cluster sample was applied) were given priority in the processes of checking and scrutiny.

The ability for providing of the first results only one year after completing of the Census is a very important effort to be able to have the data dissemination to users timely. Although the sample results are reliable in general, it is careful when explaining and using the census results because of some sampling errors. It is the same in the case of some issued figures in analysis have been rounded until several thousands units or only are represented on graphic/map while sampling errors is included in small value data. Of course the results from the complete census will have no sampling errors. The indicators on fertility and mortality, because of being collected from sample census, are allowed to tabulate and expanded to the provincial level.

Two years after fieldwork completion, processing of the 1999 census data is finished. The completed results are tabulated for all level of administrative units (country, province, district and commune).

Some depth analysis monographs are also prepared on the following topics such as marriage, fertility, mortality, migration, urbanization and employment. Based on the census figures, population projection is made for province level.

Results of the 1999 census have been made and disseminated in the following media:

Publication:

- i. Hand counts
- ii. Sample results
- iii. Completed results
- iv. 3 analysis monograph
- v. Population projection.
- vi. Socioeconomic atlas of Vietnam based on the 1999 Population and Housing Census titled “A depiction of the 1999 Population and Housing Census was created.

Electronic media:

- i. The CD ROM for data and result of the 3% sample is produced and can be seen as one of the electronic product to meet the information needs of users.

There are two main modules in the CD ROM:

First module: Keep micro data of 3% sample (from household and individual questionnaires) in IMPS format and layout about each item can be found in the data dictionary.

In addition to the data file, the CD ROM also provides some modules of IMPS (Integrated Microcomputer Processing System) and its applications as tool for producing tables and thematic map.

In IMPS, Crosstab is a module that allows users to make their own statistics table. It is located on the CD ROM containing the micro data and data dictionary. MapView is a module that helps users to create electronic maps on population information. It can be seen a simple electronic population ATLAS.

Second module: Contains tools for converting the original raw data files from the IMPS environment into more common file formats such as delimited text for imports to Excel, SPSS, and Microsoft Access.

- ii. CD ROM as the same as above mentioned for each provinces. There are 61 CD ROM for 61 provinces. This CD ROM contains: micro data, ready cross-tabulated tables, tools for converting, and Crosstab module.

- iii. CR ROM “Census PopMap application”: This provides a database with 232 indicators of the 1999 census at level of whole country, province, districts and some basic mapping classes such as administrative units, transport lines etc.

iv. The publication of “Socioeconomic atlas of Vietnam” has produced in CD ROM for easy use.

### ***3. Plan for the 2009 census***

Population size of Vietnam in the 2009 census is about 87.2 million peoples living in 21 million households. Two type of questionnaire (long form and short form) would be used. Size of sample is about 15 percent.

For the 2009 census only three regional computer centers (Hanoi, Danang and Ho Chi Minh City) would be assigned for data capture using Intelligent Character Recognition technology (ICR). Each centre would be a computer network for the census data processing with scanners, PCs (workstations and servers), laser printers, data storage devices, and the Mid Volume Capture Software and Eyes and Hands for Forms software and other related equipment.

Computer editing will be done at the regional computer centers to ensure generally that processing would be undertaken only at Hanoi Computer Center. The center in Hanoi would play a leading role in setting up the programme, receiving data files from other two centers, generating tables according to the tabulation plan and managing the whole processing system.

As ICR technology would be applied for the first time in the census, General Statistics Office (GSO) would need technical and financial support from potential donors. Based on the proposed questionnaire of the 2009 census, population size and speed of scanner, it is estimated that time of data capture will be about 12 months, if 10 scanners are used. Additional equipments procured should be compatible to the new technology to be adopted as well as suitable to the system of decentralized processing. It is absolutely necessary that scanning and ICR technology have to be tested thoroughly in 2006 and 2007 to confirm that it could be adopted successfully at the census.

Some preliminary steps have already been taken in this regard. Five staff members of the GSO were sponsored by the Government for a study tour of National Statistical Office, Thailand to study the use of scanning and ICR technology for data capture in census. Another batch of five officers would be undertaking a UNFPA-funded study tour in October 2006, of the National Statistics Office, Philippines for learning from the experience of that organization in the application of this technology for their Population Census in 2000.

The GSO has been equipped with one FUJITSU scanner and scanning software from the Swedish company “Readsoft”. The GSO have undertaken test of ICR using with questionnaire of some household survey.

The GSO has prepared a list of equipment to be procured for this new venture which is given below:

Equipment	No. of equipment in			Total
	Hanoi	Danang	Ho Chi Minh City	
1. Scanner	7	3	4	14
2. Server	4	2	2	8
3. PC	70	30	40	140
4. Software for scanning (Eye's & Hand)	1	1	1	3
5. Others equipments (Printer,UPS, air Conditioner, Humidifier, Voltage stabilizer, Networking equipment.)				

- 10 PCs will be used to verify data input from one scanner;
- One sever will be used for data backup from 15 to 20 PCs;
- One server will be used in Hanoi center to receive data from other two centers and to store whole set of census data;
- One printer will be used for printing working reports for a group from 15 to 20 PCs and one server.
- Three high speed printers will be used in Hanoi center to print the census results.
- One air conditioner for a group 20 PCs and one server, is included as part of the unit though it is not a DP equipment.

The vendor company would be providing the basic training to the data processing personnel of GSO. There would also be consultancy missions of experts in the initial stages. It is expected that trained staff of GSO would be able to successfully implement the new technology in data capture.