

Distr.
GENERAL

WP.19
30 March 2010

ENGLISH ONLY

**UNITED NATIONS ECONOMIC COMMISSION
FOR EUROPE (UNECE)
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE EUROPEAN
UNION (EUROSTAT)**

**ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE**

Meeting on the Management of Statistical Information Systems (MSIS 2010)
(Daejeon, Republic of Korea, 26-29 April 2010)

Topic (iv): Innovation and related issues including census systems

**IT Strategy for 2008 General Population Census of Cambodia at National Institute of
Statistics, Ministry of Planning in Cambodia**

Prepared by Mr. Meng Kimhor, National Institute of Statistics, Cambodia

I. Introduction

1. The National Institute of Statistics (NIS), which is part of the Ministry of Planning, is the focal point on statistical matters in Cambodia. The NIS compiles and consolidates statistics provided by decentralized offices and also collects primary data through households and establishment surveys and population, agricultural and economic censuses.
2. Cambodia has a decentralized statistical structure. There are statistical bureaus and sections within Planning and Statistics departments of the various Ministries and in planning and statistical units in the provinces and districts.
3. The IT operation in the National Institute of Statistics, Ministry of Planning has evolved with the support of many development partners, financially and technically, since 1993 up to date for the Surveys and Censuses.
4. Commencing from 1998, the Royal Government of Cambodia is committed to conducting a general population census every ten years in accordance with the U.N. recommendations and the Statistical Law of Cambodia. Such a census would include some aspects of housing census also. Accordingly the 2008 General Population Census of Cambodia was conducted on a de facto basis with reference to 3 March, 2008, exactly ten years after the previous census was held. This census was marked the second census since Cambodia become a democratic country and the first of twenty first century.
5. The IT strategy for 2008 General Population Census comprises a large array of activities, ranging from pre-census tasks; such as automated mapping, GIS development and EA database design, to post-census activities; like manual coding and editing of questionnaires, data entry, data verification, computer editing, tabulation, the development of various census dissemination products (Table Retrieval System, Community

Profile System, Population Database and Mapping Application based on 2008 Population Census) and maintenance of the website.

II. Strategic Objectives of IT Operation for 2008 General Population Census of Cambodia

6. The IT strategy for 2008 General Population Census of Cambodia was designed based on:

- (a) Data Processing Plan
- (b) Tabulation and Data Analysis
- (c) The dissemination Plan

A. Data Processing Plan

7. These are described in the following paragraphs.

(a) Pre-census data processing activities: The reference date of the General Population Census of Cambodia 2008 will be 3 March. A number of preparatory activities need to be completed before this date to ensure the availability of Enumeration Area (E.A) maps and databases, to prepare the data processing site, to procure the necessary equipment, and to develop the required training manuals, coding schemes, and software application.

(b) Post-census data processing activities: These tasks are mainly related to the actual processing of the numerous census questionnaires and related forms, to data capture, computer editing and the tabulation, to data file consolidations and backup, and to development of various census dissemination products.

8. The census data processing division of the NIS is responsible for check-in of census documents, manual editing and coding of question, data entry, computer editing and tabulation, and the generation of products like the population database and maintenance website. The data processing division has established its editing, coding and data entry teams in the different places.

9. In the check-in phase, more than 2.8 millions have been scan by using barcode reader to make sure all census questionnaire have received at NIS Census Office or not.

10. The CSPro package is used for data entry, computer editing as well as tabulation. The staff of the data processing division has been trained by the data processing consultant in using this software.

11. The manual coding and editing of the census schedules were organized in three stages. First the enumerator's summary statements were scrutinized and edited followed by data entry. This provided the basis for the provisional census results (population by sex at national and provincial levels) released through the earlier publication. The second stage related to coding and editing of the main census questionnaire (Form B). This was be followed by editing of the House list (Form A) at the third stage. The computerized census data were used to generate a number of priority tables according to an approved Tabulation Plan. This tabulation was completed in June 2009.

B. Tabulation and Analysis of Data

12. The experience gained in the 1998 Census stood the National Institute of Statistics (NIS) in good stead in planning the analysis and dissemination of the data of the present census. In consultation with data users NIS decided to produce 70 basic priority tables for the 2008 Census. Taking into account the different levels of production of these tables, the total number of priority tables in the 2008 census works out to 73,151. This is a huge increase (more than five times) from the 11,513 priority tables produced during the 1998 Census. These tables cover most of the topics included in the census questionnaire and their cross classification should satisfy most of the requirements of all sections of data users.

13. The production of priority tables are followed by the preparation of additional tables called supplementary tables proposed by the data users, and other tables produced in the course of in-depth analysis.

14. The analysis of census data by NIS will be mainly confined to priority tables and additional tables that may be produced for analysis. However census analysis is not confined to the Government alone. The universities, scholars, demographers, institutions and data users may also use and analyze the data in their own way according to their need and interest. They will be encouraged to contact the census Data Users' Service Center (DUSC) for any additional information, tables or reports. The vastness of the data and the various cross classifications and tabulations that could be made with the census tables and the micro data permit unlimited scope for the census analysis exercise.

C. The dissemination plan

15. The findings of the 1998 Census data were successfully disseminated both through electronic products and printed reports and by way of dissemination seminars and workshops at the national and provincial levels. The CDs released, each for priority tables, aggregated commune database, POPMAP applications, and WinR+ Population Database were well received and used by the line Ministries, International Agencies, NGOs, planning offices in the provinces and districts, the universities, individual and institutional scholars and researchers, teachers and students, and other data users. Another important dissemination product developed was a web site with census background information, key census results, and a request page such that distant data user can demand for further detailed census information. Census tables were also supplied as demanded by data users from time to time. DUSC has also been servicing data users on an on-going basis. All these measures promoted large scale utilization of census data by line Ministries and even government departments like Provincial Planning Offices. Training workshops were conducted at national and provincial levels on retrieving the census data from electronic dissemination products and data utilization. This is continued in 2008 Census also.

(a) 2008 Population Census Dissemination by print media

16. The reports mentioned under the analysis plan are being printed and published. Also published were handy data sheets and brochures containing important indicators as derived from the census analysis with suitable and attractive illustrations. Wall maps/charts and census thematic atlases were also produced.

17. Apart from reports and maps, census priority tables on each topic would also be published. In the 1998 Census, Tables at National and Provisional levels were published. In the context of growing literacy and educational levels among the people of Cambodia, it may be useful to make available select abridged tables also at district/commune levels. This may be useful for local planning and for those who may not have access to computer facilities and consequently may not be able to avail of the census electronic products.

(b) The 2008 Census Electronic Dissemination products

18. The fast-growing uses of computer and its networking call for wider and deeper electronic dissemination products with user-friendly interface, and efficient retrieval and manipulation functionality. The proposed electronic dissemination is classified into two main categories, 1. Off-line electronic dissemination products and 2 On-line electronic dissemination products:

(i) Off-line electronic dissemination products

Off-line electronic dissemination products will be mainly in the form of CD-ROM. The project plans to produce a variety of electronic dissemination products based on CD ROMs. These include: a Table Retrieval System, a Community Profile System, a population database built on census micro data, a thematic mapping application and Cam Info updates. About 1000 CD-ROMs may be produced initially.

Table Retrieval System

The Table Retrieval System stores the large number of census tables onto a single CD ROM and facilitates easy retrieval of selected tables. A user-friendly interface will be available to select multiple tables for multiple geographical areas. The application will also include a table viewer that allows exporting the tables or table cells to Excel spreadsheets for further manipulation and analysis.

Community Profile System

This dissemination product is based on an indicator database consisting of aggregated counts, rates and ratios for all possible geographical levels, i.e. the Country (total, urban and rural), Provinces (total, urban and rural), Districts, Communes, and Villages. A variety of indicators may be considered, such as: distribution by age groups and sex, household types, median age, dependency ratio, singulate mean age at marriage, literacy rate, employment rate, educational attainment, proportions of migrants, etc. The application will allow aggregation of selected areas and will include functionality to present some of the key indicators in a graphical format and to export the profile to Excel format.

Population database

Population databases are highly recommended as they greatly expand the usability and enhance the dissemination of census data. Databases on micro-data (individual records) permit retrieval of data at any level of detail. They are ideal tools to produce small-area statistics. However, the issue of confidentiality will have to be considered and may require re-coding of some of the variables to a higher level.

The interface for the population database will be the REDATAM+SP package. REDATAM stands for Retrieval of Data for small Areas by Microcomputer. This package enables data users to easily derive information from the database, including new variables, tabulations and other outputs. All this can be achieved via graphical windows and without the assistance of a programmer. The software also facilitates the processing of external databases in one of the common formats such as dBase and Excel. A Data Dictionary, describing in detail the structure of the database, will be included with the product.

Mapping application

Mapping and graphing databases also greatly improve the effectiveness of census dissemination as trends and patterns of the larger area, and distinct boundary and characteristics are more easily detected when displayed on maps. The project will make available map layers for the Country, Provinces, Districts, Commune and, if possible, Villages. Statistical databases consisting of aggregated count, ratios and rates for all geographical levels will complement these map layers. The layers will be in ESRI shape file and MapInfo table file formats. Data users are expected to acquire a copy of their preferred GIS software.

Cam Info Updates

CamInfo is the national adaptation of Dev Info software, a global initiative funded by the United Nations. Cam Info is Cambodia's Socio-Economic and demographic Indicator Database, providing a one-stop user-friendly computer program for storage, retrieval, comparison and dissemination of a wide range of indicators from different sources, including national surveys, censuses and administrative systems. Cam Info also allows the user to retrieve and compare indicator data values across multiple time periods, geographic levels, and other sub-group disaggregations. Data presentation is possible with tables, graphs and maps.

The National Institute of Statistics has selected indicators from the priority tables of the 2008 Census based on the perceived needs of users, for incorporation into the CamInfo regional updates. The online user interface is the same as the CamInfo updates CD-ROM. This will allow users both inside and outside of the country to access data using the Internet without installing the CD. A DevInfo workshop was held in February 2010 when CamInfo of 2008 Census was released.

The feasibility of using the recently introduced census dissemination tool called UN Census Info would also be explored.

(ii) On-line census dissemination

The National Institute of Statistics maintains a web site with information on population censuses, the results of various types of surveys, periodical publication, etc. The URL is www.nis.gov.kh. Most of the information available on this web site is in a static format. For the on-line dissemination of the results of the General Population Census of Cambodia 2008 a more dynamic approach is envisaged. Priority tables and analysis outputs are suggested to be available on-line for wider and distant accesses. The possibilities of SQL database querying and on-line mapping will be explored. To abide by the statistics law on keeping confidentiality of respondents, however, security, authentication, recoding and aggregation shall be closely observed.

(iii) Other Electronic products

As there is a vast scope to expand this type of dissemination in modern times, more electronic products may also be utilized.

(c) Dissemination Workshops

19. Seminars for the presentation of census results and workshops to train planners in the line Ministries and other data users may be conducted in the course of 2009-10. Such seminars/workshops will be held in every province/district so as to benefit participants down to the village level. They may be held in Phnom Penh and provinces once in 2009 closely following the release of final results in September 2009 and again in 2010 at the national, provincial and district levels after the provincial reports are released.

20. The purpose of the dissemination workshops is to provide census data to planners, administrators and researchers at province, district and commune levels. Such direct interactions between the producers of census data and its users would enable the latter to understand the impact of population growth on welfare measures undertaken by the Government and also help monitor the progress made in the various development programmes of Cambodia.

III. Constraints

21. There were many constraints met during pre-census and post-census activities. Most important of them are:

- (a) There was paucity of supervisory IT staff in the Census Data Processing Division. This was because trained staff left the organization to take up other jobs. However, the workload was successfully put through by existing staff under the guidance of a Data Processing Consultant.
- (b) Sometimes virus-affected computers hampered the progress of work. This problem was solved then and there though it delayed the work.
- (c) Coding of 2008 Census schedules was manually done by Editing and Coding teams. During data entry, wrong codings were detected and the schedules had to be corrected.

- (d) The 2008 Census processing showed that computer editing is very essential for maintaining consistency of data.
- (e) In Cambodia several administrative area changes were made by the Royal Government of Cambodia after the 2008 Census. This resulted in re-coding of administrative units and change of the urban-rural frame. The changeover to the new frame of administrative units created more work and needed more time.
- (f) Suitable incentive to Data Processing staff may be needed to attract and retain efficient staff.

IV. Conclusion

22. The IT strategy for 2008 General Population Census of Cambodia played an important role and served as an essential tool to successfully complete the census work. It has been mentioned that due to well planned IT strategy the 2008 census results were released in record time, namely within a few months of the completion of the 2008 Census.

23. The IT strategy covered the pre-census and post-census operations. The pre-census activities included check-in phase, site for data processing, and training. The post-census activities covered data entry, computer editing, tabulation and dissemination operations. Both activities had to be strictly monitored.

24. It would be better if lessons are drawn from the constraints mentioned above to improve the IT in the future censuses.
