

Population and Housing Census Malaysia 2010 New Approaches And Technological Advancements

Prema Letha Nair

Population and Demographic Statistics Division, Department of Statistics Malaysia,
Level 4, Block C6, Complex C, Federal Government Administrative Centre
62514 Putrajaya
e-mail: prema@stats.gov.my

Abstract The Population and Housing Census remains vital as the main source of data collection aimed at providing a comprehensive set of statistical information about the population in the country in terms of its size and spatial distribution, its demographic, social and economic characteristics as well as housing stock at a specific time reference. The census provides essential information for policy development and planning, for managing and evaluating programmes activities across a broad spectrum of sectors and for monitoring overall progress. The Census being a total coverage and complete count of the population, household and living quarters at a given time enables to provide for a wide scope of enquiry even to the extent to meet the demand for data on small areas statistics. The United Nations Statistical Commission launched the current 2010 World Population and Housing Census Programme covering a 10 year period from 2005-2014. It urges member states to carry a population and housing at least once over the above said period. Malaysia, launched the fifth decennial Population and Housing Census in July 2010 which provides a wealth of information on the population and household with demographic and socio-economic characteristics as well as housing stock. The 2010 Census round witnessed increased interest and use of improved technologies and new approaches in the different phases of the census operation. The emerging improved technologies are transforming the way the Census are conducted in terms of operation management, quality assurance, data capture, mapping (GIS, GPS), data processing and storage. Among the new approaches used during the Census are the use of internet based questionnaire for data collection, alternative publicity strategies being implemented and the web-based data dissemination. Even small improvements in census technology can result in important gains in the quality and cost-effectiveness of the whole census operation and still maintain the confidentiality of data. This paper outlines how the Department of Statistics Malaysia adopted new approaches and technologies during the 2010 Census. Highlights will also be given on the impact of the new technologies in the implementation of Census and the way forwardness.

Keywords Population and Housing Census

2010 Mathematics Subject Classification 62P25

1 Introduction

A population and housing census represents the principal means of data collection aimed at providing a comprehensive set of statistical information on population and housing. The Department of Statistics Malaysia (DOSM) is the national statistical office in Malaysia. Since the introduction of the Census Act 1960, DOSM has been entrusted by the Federal Government to be responsible in the undertaking of the Population and Housing Censuses for the years 1970, 1980, 1991, 2000 and recently the 2010 Census apart from its responsibility of collecting, interpreting and disseminating statistics in other fields. The Census covers a wide range of key socio-economic characteristics on persons, households and living quarters. Census is the only source of data where population data at micro level is made available compared to other statistical collection. It provides benchmark data for all demographic, social and labour force statistics as well as a basis for the demarcation of electoral

constituencies boundaries and distribution of Federal Government funds. Data from the population census are used as inputs for policy planning, formulation and implementation of programmes.

This paper examines the new approaches and technological advancements that was used by DOSM in undertaking the Census 2010 and the way forwardness to adopt and further enhance strategies for the next census. The paper will discuss on the following:

- (a) Advent of technology advancement in Mapping;
- (b) Adoption of new technology in data processing ;
- (c) New methods in Census publicity and data dissemination;
- (d) Monitoring the Census field operations;
- (e) Multi-modal approach in Census taking; and
- (f) Development of central repository population database;

2 Advent of Technology Advancement in Mapping

Census processes entail three stages; namely pre-census, census and post census. Maps are important geographical tools to assist in data collection and monitoring census activities. Recognising this, DOSM in the Census operation and coverage for mapping implemented a two tier statistical category which has various geo-statistical level as identified below:

- (a) **Administrative areas** which are gazetted areas used in the census, such as administrative district (AD), mukim, sub-district and local authority area (LAA); and
- (b) **Statistical areas** which are created and defined by DOSM for census operation such as census district¹ (CD), census circle² (CC) and the lowest sub-division being the enumeration block³ (EB).

Mapping being an activity of pre-census has been an integral part of census taking for a long time. The whole country was subdivided into EB's to ensure full coverage and facilitate in census operations. The advent of computers as information processing tools and the development of Geographic Information System (GIS) have measurably assisted the geography discipline. DOSM realising the potentials of GIS in census work had started the development of GIS in 1989. Initial studies were conducted pertaining into the viability and availability of hardware and software. A consultant from the US Bureau of Census was engaged to assist DOSM in this new technology development.

Figure 1 EB sketch map for 2000 Census



¹ CD consists of 100-120 enumeration block. Used by Assistant Commissioner & District Superintendent to plan and control census operations

² CC consist of 7 enumeration blocks which is used by the Supervisor for operational control of the designated area of the enumerator

³ EB is an assigned coverage area of an enumerator covering a range of 80-120 living quarters (LQ).

Figure 2 Digital EB map produce using GIS for 2010 Census

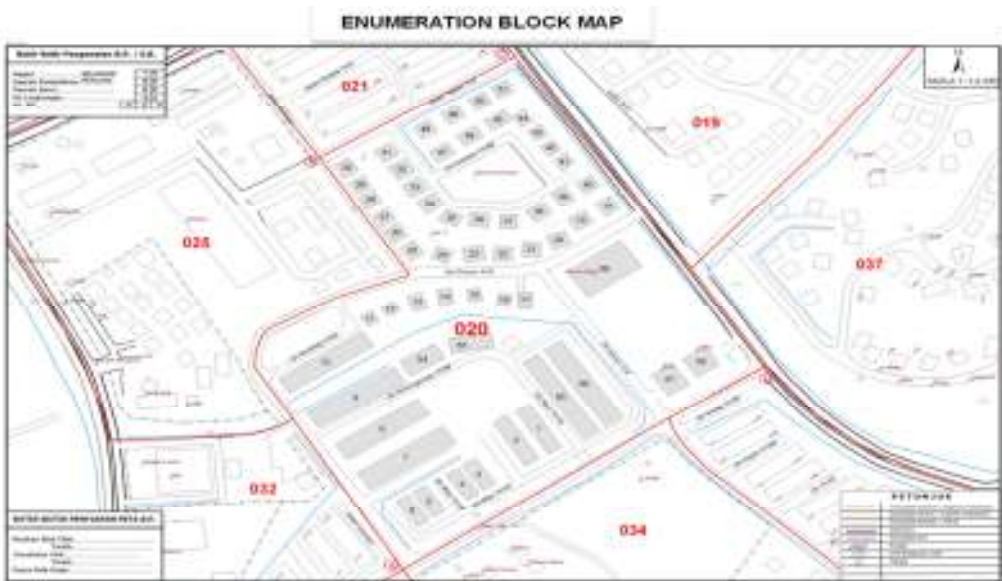
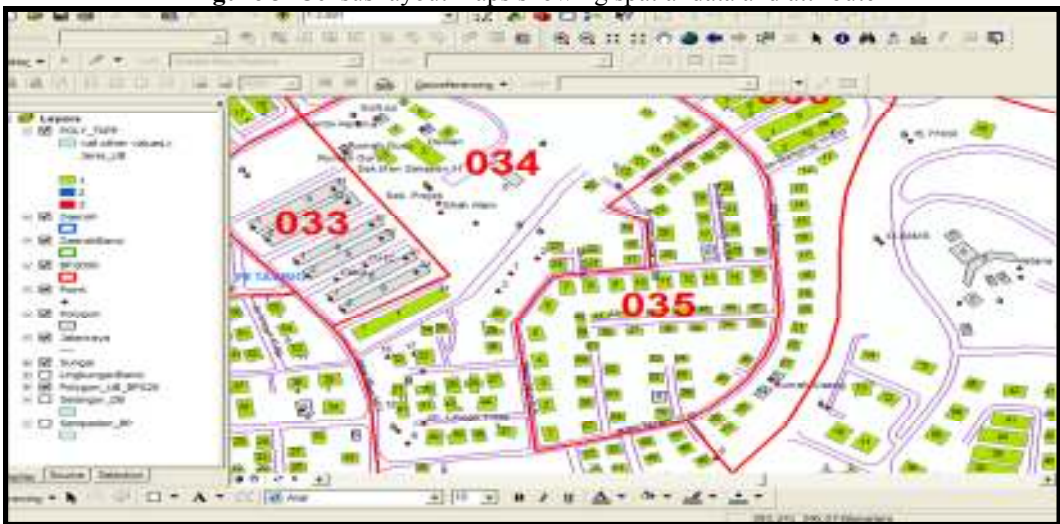


Figure 3 Census layout maps showing spatial data and attribute



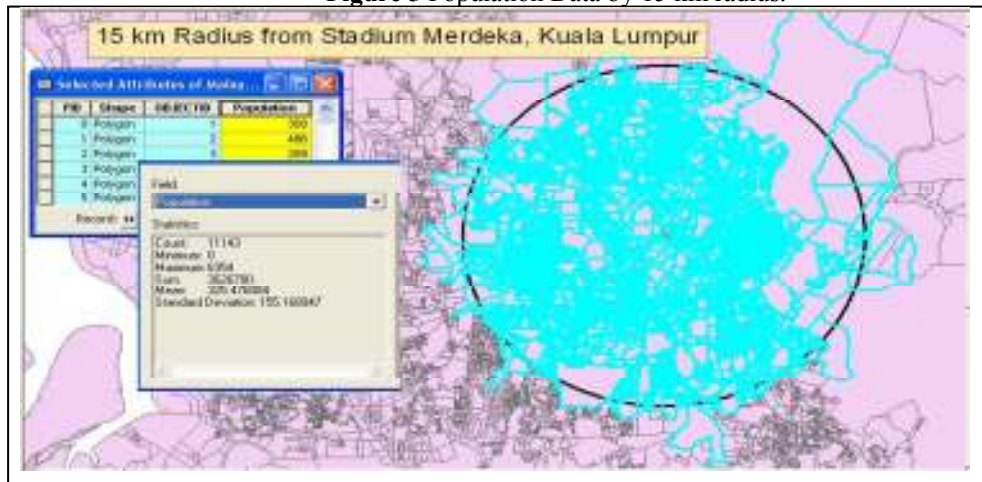
Spatial data joined with attribute

NO	WILAYAH	KECAMATAN	KELURAHAN	KEBAYAKARAN	ALAMAT	NO HP	STATUS	REMARKS		
NO	WILAYAH	KECAMATAN	KELURAHAN	KEBAYAKARAN	ALAMAT	NO HP	STATUS	REMARKS		
1043005	003	2155	Sykt AA	3A &	TKT 1 & 2			JUN TEWSU AMIRIAN Z	SEKSYEN 9	SHAH ALA
1043005	003	2155	Sykt BB	3A &	TKT 1 & 2			JUN TEWSU AMIRIAN Z	SEKSYEN 9	SHAH ALA
1043005	005	2298	Sykt CC	05-01	TKT 5	PLAZA MASALAH		JUN TEWSU AMIRIAN Z	SEKSYEN 9	SHAH ALA
1043005	005	2298	Sykt DD	05-01	TKT 5	PLAZA MASALAH		JUN TEWSU AMIRIAN Z	SEKSYEN 9	SHAH ALA
1043005	005	2201	Sykt EE	LOT-4	TKT 1	SHAH ALAM MALL		JUN TEWSU AMIRIAN Z	SEKSYEN 9	SHAH ALA
1043005	005	2201	Sykt FF	LOT-4	TKT 1	SHAH ALAM MALL		JUN TEWSU AMIRIAN Z	SEKSYEN 9	SHAH ALA

Figure 4 Population data (aggregated) using buffering process.



Figure 5 Population Data by 15 km radius.



The actual GIS work started in 1993 with the undertaking of digitizing the 1991 Census EB boundaries where the information was kept as a cartographic data base. Over the years of 1993-2000 the preparation of census maps took a turn from a hand drawn manual process to the digital format as in Figure 1 and Figure 2. The enhancement of GIS will benefit DOSM to produce EB digital formats. Users of EB digital maps will be able to find, manage, retrieve and restore the EB digital maps in a much easier manner as all the data is stored in a GIS database. DOSM expanded the GIS system which was once concentrated at the head office to the state offices. Extensive training was provided for the staff and all the state offices were equipped with GIS facilities. As a result monitoring and controlling of all updated EB maps (spatial landmarks and features) and the attribute data becomes more efficient using GIS as in Figure 3. The different levels for both spatial and attribute data was further built-up for the 2010 Census. The spatial data for Census 2010 saw an improvement in digital format where layers for building unit and living quarters was included apart from the spatial data of state boundary, AD/Mukim, CD and EB maps. In the case for the attribute data the additional specification levels are living quarters and household/individual. These improvements using the GIS application added value to DOSM products where most the users need for small area statistics were met. In addition the population demographics in terms of aggregated data through the buffering process and aggregated based on the EB number can generate data by radius as in Figure 4 and 5.

Moving forward, DOSM will explore areas in GIS innovations methods of generating, presentation and dissemination of data. Enhancement in human capital skills is essential to enable ever increasing data especially of small area statistics.

3 Adoption of New Technology in Data Processing

Among the government agencies, DOSM is one of the leading pioneer agencies to implement the use of technology in their work. The evolution of ICT development in DOSM witnessed the use of the Optical Mark Recognition (OMR) technology during the 1970 and 1980 Censuses. In order to enhance the usage of personal computers among the personnel as well as the strategy towards decentralisation of data processing DOSM opted for the manual data entry system for the 1991 and 2000 Censuses. In the 2010 Census, for the first time DOSM adopted scanning as a new data capture method. The ever increasing volume of questionnaires and with the challenge in complying to standards and requirements several rounds of testing using the Intelligent Character Recognition (ICR) technology were carried to gauge the problems, advantages and disadvantages were identified prior to the capture of census data. The ICR machines were located in seven processing centres. Based on reviews from other countries census taking and surveys conducted by DOSM (manufacturing and distribute trade surveys) that have used the ICR technology the illegible handwritten entries was among the main problems faced. Specialised computer software was used to interpret the handwriting on the respective images of the census forms and transforms them to into single computer data file. The time taken in processing the data was reduced and hence reduces the intake of staff needed when compared to manual driven process used in the previous censuses.

Figure 6 Image for the computer assisted coding on screen.



The issue of computer assisted coding (CAC) has been the subject of considerable research, [1]-[2]. Coding is done for some variables especially where answers are provided in free text example occupation and industry and a drop-down list is provided to match the best fit code to the description of given free text as shown in Figure 6.

CAC provides a friendlier and easier search function for coding of occupation and industry compared to manual coding which requires time and effort for identifying the right codes. The pre CAC edit specification allows the basic checks to be done electronically. Once the forms passed pre CAC edits, further onscreen and online coding processes for education and migration is conducted. The CAC improves coding activities by enhancing the quality of operations reducing coding errors and speeding up the coding process.

The chances of finding respondents at home during working hours have become smaller due to modern lifestyles and smaller household size. Around the world, the wide access of electronic mail has increased over the years and we witness the use of the Internet as a medium to complete the census questionnaires. DOSM for the first time introduced this approach with the aim of leveraging on the internet savvy community. The e-Census approach was found to be more effective to capture respondents mainly in large urban areas and the gated communities. With the ever increasing emphasis for individual privacy, the increase accessibility of internet, it was the right time for the Department to consider options in its approach to census taking. Hence the outcome was to capitalise on the existing situations such cost to hire enumerators, the need of privacy and the unwillingness to divulge individual information, the Department decided to use the internet approach.

Internet (e-Census) technology represents the greatest opportunity to increase the efficiency and value of a population census, but at the same time present the greatest challenges. Strict security measures at various stages of process were imposed to ensure confidentiality and safety of the information provided by respondents. Username/passwords and pin numbers were given to households during the visits by enumerators. Notification of e-Census submission was automatically sent to the field supervisors via Short Message Service (SMS) for the field monitoring. Built-in checks were installed to check for consistency and quality of data. Also measures were taken to remind respondents as to the need to complete the questionnaire within a specified time through the process of status tracking. Through this monitoring process if the time frame in answering the e-census questionnaire is not adhered, steps are taken to visit and interview the household.

4 New Methods in Census Publicity and Data Dissemination

Census is the largest statistical project conducted by DOSM. Census publicity during census operations and dissemination of the census findings also means increased awareness of the Department among the population at large. The need to provide awareness, understanding and information as to the purpose and importance of census taking was the key role function of the Department. Assistance was sought from other government and non-governmental agencies and the public media at large for a wide national coverage to ensure awareness and to gain cooperation from the general public in providing census information. Several strategies were formulated and implemented to ensure the public awareness of census. Within the Department a census logo and song competition was launched. Publicity campaign, broadcast/announcement on Radio/TV, advertisement, minister's launching of the Census, full media coverage, press conference, film trailer Census 2010, census brochures and pamphlets as well as publicity through electronic media mySMS and internet. Even a statement was published in our electricity and water bills to further enhance the awareness of the Census. During the census a call centre was set up to tackle issues such as living quarters that were missed out, query on the census enumerators, how to complete the census questionnaire through internet and others. A social network such as the Facebook site was created to handle issues or queries raised by the public.

In the previous censuses the primary means of data dissemination was through printed publications. The development in the IT industry which enables desktop publishing has revolutionized the process. The challenge to most of the statistical offices is considerable. In the earlier censuses conducted by DOSM for the years of 1970, 1980 and 1991 a publication has to undergo several rounds of checks before it can be given for printing as the numbers to be printed was large. Since

Census 2000 and 2010 DOSM any form of publication must adhere to a strict calendar of electronic release. Designing a dissemination strategy has not become any simpler. Users always want the data sooner and expect statistics to make full use of new media yet would demand in paper publication format. To meet the demands of users the latest census publication release of the 2010 Census is upload in the Census Portal website where all information pertaining to Census is made available. The basic request for data in CD-ROM, specific cross-tabulation data request with maps or no maps, e-mails, mails, telephone request and a 2% micro data by researchers is always met.

5 Monitoring the Census Field Operations

The adoption of new methods and technology, witnessed for the first time an on-line monitoring process of the progress in the field enumeration activity being implemented during the Census 2010 round known as the e-RKL system. The system was developed in-house where it uses the Hypertext Pre-processor (PHP) for the programming language, software Dreamweaver and MySQL as data base. Once an enumerator completes the assigned EB, the supervisor checks and sends the report of the preliminary figures of population, household and living quarters to the district office. The District Superintendent (DS) upon receipt of the report keys in the figures and sends it to the Assistant Commissioner and the Deputy Commissioner who like the DS will be able to control and monitor all census operations on the field. The operation centre in headquarters as well as the processing at state level was able to view the progress as well as bottleneck areas. This monitoring allows the census officers to strategically plan, make assignments, identify problem areas and implement quick remedial actions. It is cost efficient which is in line to the concept of green ICT and also enables in the preparation of the Preliminary Count Report at administrative district and state level. The processed data at the processing centres will be transmitted online to head office for tabulation processes. Data mining facilities expedites the production of census output and releases of Census product compared to traditional mainframe processing.

6 Multi-modal Approach in Census Taking

In the 2010 Census a multi-modal data collection strategy, comprising the traditional face to face interview in both urban and rural areas; self-enumeration paper questionnaire which was drop-off and pick-up by the enumerator and for the first time in census data collection a self enumeration questionnaire via internet was implemented to target urban areas. Respondents or households for the first time had the choice in selecting which method most preferred thus increasing the response rates compared in 2000 Census.

The enumeration period was from 6th July to 22nd August 2010. The time frame to complete the census was lengthened to 6 weeks compared to 2 weeks in 2000 Census. In field work operations unlike 2000 Census where an enumerator was given a single EB for a period of 2 weeks, in the 2010 Census an enumerator was given 3 EBs which had to be completed within 6 weeks. While the supervisors was in charge of monitoring 7 enumerators (21 EBs) compared to 7 enumerators (7 EBs) in 2000 Census.

7 Development of Central Repository Population Database

In view of the Department's ICT Strategic Plan, the Census metadata, census data and GIS have been integrated in the National Enterprise-Wide Statistical System (NEWSS). It is a platform for an integrated statistical system for collection, processing and dissemination of statistics. The Census data for the years of 1991, 2000 and 2010 will be kept under NEWSS repository which will harness future data request.

8 Conclusion

Leveraging on the success and lessons learned from the 2010 Census, DOSM has to look into the possibility of conducting the next population and housing census in 2020. The need for continuous census management will help DOSM to adopt new approaches and technology advancement to produce more detailed, high-quality, and timely statistics with limited resources to keep abreast with the customer/user expectation.

The 2010 Census, DOSM has enhanced the usage of ICT in the mapping process using GIS, data collection via e-Census and data processing via ICR. The creation of in-house systems aided in the monitoring the stages of progress in field operations as well as assisted in the coding process. Built-in checks and edit specifications enabled to produce quality and timely data. To cater for the changes in social and demographic pattern of the population new approaches to census taking was introduced. The task in the conduct of the 2010 Census can be summarized as “**making Census easier and results more useful**”. DOSM has managed to get a higher response rate and at the same time managed to meet her data requirements as far as possible. The 2010 Census has given the opportunity for the Department to showcase the adoption of IT advances and innovative measures in Census-taking and enhance its professional image, integrity and reputation as a progressive and sole provider of statistics in the country. Keeping this in view, DOSM in the 2020 Census will continue to explore new challenges in IT innovations to make data collection and processing easier.

References

- [1] Dopita, Patricia. *Population Census Evaluation, 1996 Census Data Quality: Occupation*. Canberra, Australian Bureau of Statistics. 1999.
- [2] Blum, Olivia. Editing and Coding Module. In *New Census Technologies: The Isreali Experience. Proceedings of the Euro –Med Workshop, March 1997*.