

# Thailand experience in multi-modal data capturing for the 2010 Population and Housing Census

Wongsaroje Malee

*ICT Center, National Statistical Office Thailand*

*The Government Complex, Chaeng Wattana Road*

*Bangkok 10210, Thailand*

*E-mail: mwongs@nso.go.th*

Wasintham Somsri

*ICT Center, National Statistical Office Thailand*

*The Government Complex, Chaeng Wattana Road*

*Bangkok 10210, Thailand*

*E-mail: somsri\_w@nso.go.th*

Thailand's first Population Census was conducted in 1909. Since 1960, NSO has carried out Population Census every 10 years. The latest Population and Housing Census was conducted in April 2000 and ICR technology was first implemented for data capturing. The Population and housing census 2010 is a 100th anniversary for the census of Thailand. Multi-modal collection methods are used for data collection in order to provide more alternatives for population to participate in the Population and Housing Census. Internet and telephone interviewing center (TIC) are implemented for the first time. ICR technology is still used as the core tools for data capturing, while newly Internet and TIC are alternative methods for people who are hard-to-reach by interviewers or people who are more comfortably to give information through internet of telephone. Moreover, drop questionnaires are requested to send back to the offices by post in the case interviewers met nobody during field work enumeration. Thailand has used the multi-modal collection methods for the first time and careful evaluation are carried after the implementation during September 2010. Various lessons are learnt and shared as our experiences.

## I. Historical Background of Thailand's Population Census

Thailand's first population census was conducted in 1909 by the Ministry of Interior, during the reign of King Rama V. Four subsequent censuses followed in 1919, 1929, 1937 and 1947. Since 1960, the National Statistical Office (NSO) has been responsible for undertaking population censuses every ten years under the 1952 Statistical Act (revised in 1965 and 2009). In accordance with the United Nations' recommendation, that countries should undertake national censuses every the year ending with 0 (zero) for the purpose of international comparison, Thailand has conducted its census in 1970, 1980, 1990 and 2000. In 1970, the first housing census was conducted simultaneously with the population census. The tenth population and fourth housing census was carried out in April, 2000. Total population and households were 61m and 16m respectively. The latest eleventh census Population and Housing Census was carried out in September 2010. This census also marked 100 years of population and housing census in Thailand.

## II. General Preparation for the 2010 Population and Housing Census

In Thailand, the most concerned problems, which Thai National Statistical Office (TNSO) has been facing in conducting Population and Housing Census is uncooperative from some particular groups of respondents. This has caused a significant under-counted number of populations. TNSO, which is the

responsible agency for the census, has recognized the problem. Thus the strategic plans had been set up for the 2010 round of the census to address the problem as follows;

- Establish the cooperation among alliance agencies Thai NSO signed Memorandum of Understanding (MOU) with 8 agencies of both public and private sector in 2008, in order to join hands in conducting population and housing census, in term of sharing knowledge and human resources.
- Public relation and campaign of the project in order to advocate public of the population and housing census. The effectiveness of the media campaigns had planned to promote population and housing census to public. Several media and channels are used such as announcement in newspapers, broadcast by community radio stations and televisions, poster drawing competition, census logo and slogan contest include distributing promotional calendar of Census 2010 and issuing memorable stamps. A mascot for field work operations was introduced and widely known by general public.
- Establishing community network: to create awareness about conducting PHC and interest to communities, NSO did the Pilot project of 2010 Population and Housing Census in May 2008 in Phitsanulok Province, of totally 84 enumeration districts (ED) in 3 piloted districts, 6 sub districts. Prior to the pilot test, NSO had called for a meeting among local community authorities in order to better an understanding of conducting the PHC as well as to ask for their cooperation.
- The use of multi modal data collection including strategic plans for preparing proper instrument and human resources. The special treatments in particular areas, such as high-income-housing area, high-rise buildings are designed. Multi modal data capturing are Internet and Telephone Interview Centre (TIC) are in addition to Intelligent Character Recognition (ICR) questionnaire forms.

Despite of various preparatory activities, the NSO faced with unexpected event, political turmoil, before commencing the field work. The 2010 Population and Housing Census was postponed for 2 months from July 20 September 2010 and was extended for another month, as periods of rains have hindered officials' efforts during information collection.

### **III. Data Capturing Preparation for the 2010 Population and Housing Census**

For the 1960, 1970, 1980 and 1990 Censuses, the main method of data collection was field interview. After the data collection in all provinces was completed, questionnaires were sent to the central office in Bangkok. Manual editing, keyboard data entry, and other steps of data processing, including tabulation were then carried out. The quality and timeliness of the data provided by the census can always be improved in order to meet the needs of various users. The 2000 Population and Housing Census in Thailand adopted Image Scanning and ICR System for data capture. This new technology decreased the number of staff members for data entry and timeliness of data entry.

Lessons learnt from using ICR system during the 2000 Census were used as an input for preparatory of the 2011 Census. Besides, special particular areas, such as high-income-housing area, high-rise buildings are recognized during planning activity. Moreover, Information and Communication Technology (ICT), in particular became more common used in general public. Taking those factors into account, the 2010 Population and Housing Census adopted the multi modal data collection. Four methods of data collection are following:

- Face to face interviewing by enumerator, as the major method of data collection, filled out questionnaires were scanned and data were captured by Intelligent Character Recognition (ICR) system;
- Filling out a self-administered questionnaire and mailing this back to the NSO head office (Drop off and mail-back);
- Filling out the data on the Internet (Self Interview); as an alternative for people who used internet and willing to provide data on their own convenience;
- Telephone interviewing through Telephone Interview Center (TIC), as an alternative for people who were willing to provide data on their own convenience.

The period of data collection is 1-31 September 2010 with census reference date of 1 September 2010. There are approximately 70,000 enumerators working for the whole country. The Internet Web Site and Telephone Interview Center commenced about the same time. The data collection was carried out for one month period. The process has been well received by people nationwide, judging from cooperation of the public throughout the country. However, rains have hindered the data-collection process, which originally is scheduled to be completed in September, 2010. The Office has thus decided to extend the data-collection period for another month, till the end of October 2010.

Following sections explain in details data capturing methods adopted in the 2010 Population and housing census. Lessons learnt from ICR system, Internet and Telephone Interviewing Center are focused.

#### **A. Scanning of the questionnaire data using ICR (Intelligent Character Recognition)**

For the 2000 Census, the NSO was firstly used Intelligence Character Recognition (ICR) System. to process the Population and Housing Census questionnaires by scanning the 16 million households (16 million forms) which spent only 8 months to process the raw data instead of 18 months by using Key in Data System.

ICR Technology was the core tool that would be used for data capture in the 2010 Population and Housing Census. In 2008, NSO procured efficiently ICR hardware which considered efficient high speed scanners that suit for batching workload, flexible server that support large memory capacity, large capacity external storage and selected e-FLOW software with the highest recognition rate and lowest percentage error to enhance quality of data capture.

In 2010, NSO has procured the scanners and distributed to 75 Provincial Statistical Offices for serving huge questionnaires workload. The number of population in 2010 is around 19 million households. (76 million sheets: A4 paper size). The dropped color images were transferred online to central office for processing with ICR system.

At the Central Office in Bangkok, the configuration of scanning and processing system is as follows:

- Scanners: Creates drop color images and full images for Bangkok area,
- Servers: Recognition and Interpretation of captured data from the scanned images,
- Workstations: Verification and correction of uncertain or unreadable characters,
- PCs: Monitoring of the system,
- Network: Local Area Network (LAN) 1 GB/s

At the Provincial offices the configuration of scanning and processing system is as follows:

- Scanners: Creates drop color images and full images
- PC: Monitoring of the system
- Network: Government Information Network (GIN) 10 MB/s

There are both advantages and disadvantages from the decentralized design of scanning system. So far, we cannot really have solid conclusion. Due to the work is still going on. However, preliminary conclusion can be drawn as follows:

Advantages of decentralized data processing are:

- The work is done closer to the data source, where questions are more easily answered by an official,
- The costs for transporting large amounts of forms over long distances are reduced,
- Decentralized structure allows for more people to become involved in the census and gain further experience and responsibilities from the census.

Disadvantages of decentralized data processing are:

- Lack of skilled staff to manage the provincial offices,
- Local customer service and technical support agents lack of skilled to support and maintaining,
- A too tight implementation plan can force the end result to be delayed.

Some operational problems facing are examples below:

- Un-accurate size of questionnaires paper (wrong size),
- Humidity/ wet cause the paper to swell which cause error in the scanner feeders,
- Handwriting quality was below expected/ poor handwriting.

## **B. Self Interview by filling out the data on the Internet**

Internet data collection is convenience for respondents to complete the questionnaire. The objective is to decrease coverage error, providing a way to get at hard-to-enumerate households. Internet data capturing is user-friendly, reliable, unbiased and confidential.

By using the Internet-based Census form, the respondent could fill out the questionnaire him/herself and submitted it directly upon completion at <http://popcensus.nso.go.th>, or <http://www.nso.go.th>.

Quality control was assigned to control and follow up fieldwork during data collection. NSO was developed the data collection progress program for monitoring the fieldwork progress as automatically linked Internet data collection web application. Purpose of this program is to facilitate NSO director and also Provincial directors for monitoring fieldwork and quality control.

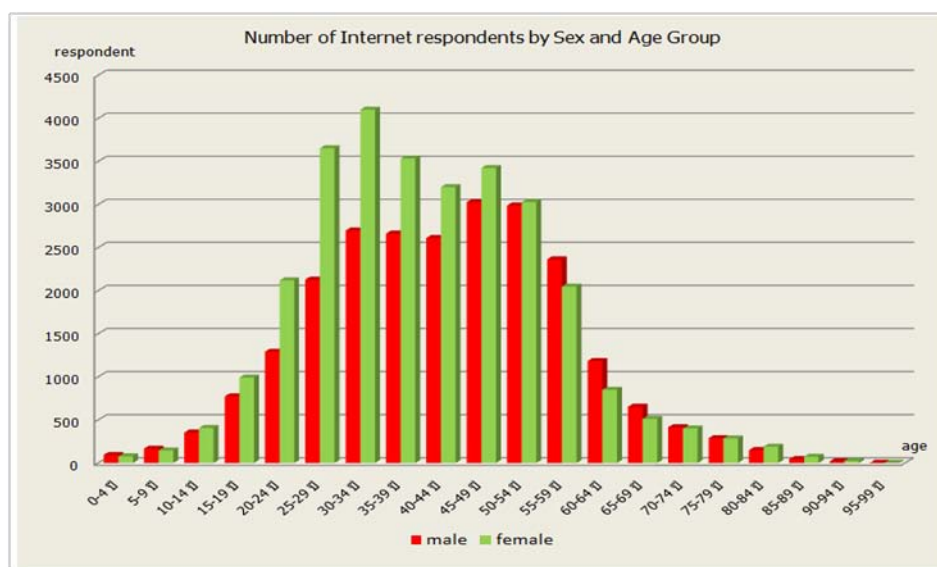
The on-line community through the Internet is an influential force in society by virtue of its speed and convenience of communication among individuals or groups. The NSO gives importance to the on-line community for its potential to help promote the National Population and Housing Census through such channels as Facebook, Twitter, and YouTube.

The benefits of internet data collection are accurate and timely data collection and reduction in the processing time. Internet provides rapid availability of clean data for statistical analysis.

Lesson learned from filling out the data on the Internet is indicated below:

- According to the Internet-based Census form that wasn't the EA boundaries detailed for the 2010 Population and Housing Census. To fulfill the Census form, the additional step would be determined to assign the EA detailed by manual checking each location of population and households with digital EA maps (GIS technology),

- The respondents suggestion about questionnaire that shouldn't ask too much details and contents should be identified the statistical definition name/words and also inform that they attempt to access through internet many times,
- The respondents didn't understand the questions and the statistics definitions/words even though the common reference with pop-up was displayed on screen.
- Due to the information and communication technology survey in 2010 of NSO, the using internet household was 2.24 million. However, the census schedule plan expected to receive response rate on Internet about 10% (0.224 million household), but the census period found that the response of Internet & TIC was 1.38 % (30,936 household) as it was lower than expected.
- Preliminary raw data extracted from the database were analyzed in order to capture basic characteristics of respondent, such as sex and age group. Below diagram shows the number of Internet respondent by sex by age group. The figure shows that female respondents by Internet are among 25-39 years of age group, while male respondents are among 30-54 years of age group. A number of female respondents are higher than male respondents.



*Diagram 1: Internet respondent by sex by age group (unpublished data)*

### C. By phone interview (Telephone Interview Center: TIC)

The respondent could also provide the Census information over the phone through Telephone Interview Center: TIC by dialing 1111 and choosing ext. 6.

The NSO has provided the computer-assisted telephone interviewing (CATI) technique that the interviewer reads the question developed on the computer screen and records the respondent's answers directly into the Web-based Data Entry.

Lessons learned from TIC are listed below:

- Some advantages for TIC are citizens can provide data at their own convenience. They could also get more information and clarification about questions if needed.
- A well organization must be informed the enumerators respondents.

- The commendation about the contact call center (TIC) line (1111 press 6) is hard to contact during day time and also have to dial many times. Regarding the connecting network on internet and telephone line extension, the system was caught down during first week of census period. Later on, the additional lines were set up to serve the call center.
- Preliminary raw data recorded by TIC operation were analyzed in order to draw a result of TIC calling lines. Below diagram show that success interviewing was lower than irrelevant calls. This is because the common number of 1111 was also used for various governmental hot-line.

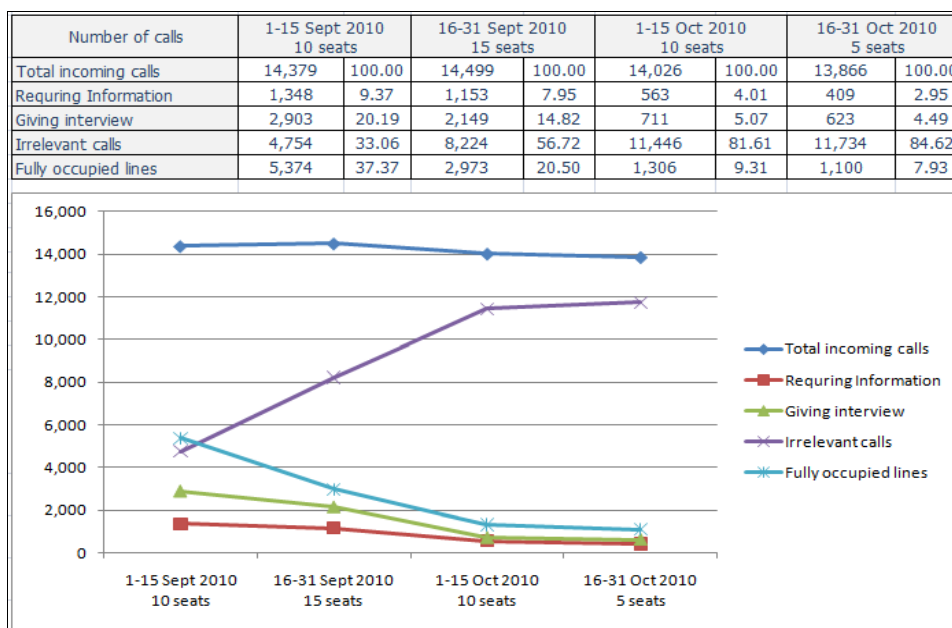


Diagram 2: TIC callers by period and type of call (unpublished data)

#### IV. Conclusion

Although the result of the 2010 Population and Housing Census is not yet released, most of data collection activities were completed. Some experiences can be preliminary summarized based on some unpublished data. ICR system is the major method for Population and Housing Census data capturing. This is the used for the second time for the Population and Housing Census. The distribution of scanning processes to 75 provincial offices is extended from the first used. Self-interviewing by Internet and Telephone Interviewing Center (TIC) were newly introduced in this 2010 Census. The internet web based data capturing were adopted as a good alternative methods for netters. However, the result was lower than expectation. TIC was aimed for respondents in a particular area such as high-income-housing area, high-rise buildings and as well as for citizens who were not ready at home for an enumerator. Surprising result has shown that irrelevant callers occupied more lines than succeeded callers for the Census interviewing. In conclusion, the newly introduced data capturing contributed very little to the 2010 Population and Housing Census with less than 5 percents.

#### REFERENCES

Jirawan Boonperm (2004), Census 2000 and its implementation in Thailand: Lessons learnt for 2010 Census, United Nations Symposium on Population and Housing Censuses, New York, USA

Jirawan Boonperm (2006), Thailand Population and Housing Census in Thailand, Paper presented at the ASEAN Meeting on the 2010 Round of Population and Housing Census, Cambodia