Enumeration, Data Processing and Dissemination

3 ENUMERATION, DATA PROCESSING AND DISSEMINATION

3.1 OVERVIEW

Census 2010 was officially launched on 12 March 2010. The sample enumeration for the Census 2010 was conducted from mid-March to end-August 2010, centred on the Census reference date of 30 June 2010. A tri-modal data collection strategy, comprising the Internet, Computer-Assisted Telephone Interviewing (CATI) and face-to-face interviews, was adopted to facilitate data collection from the households. Data collected were sent for processing where the records were edited to eliminate errors and omissions, and coded for tabulation and analyses.

For the conduct of the Census 2010 sample survey, the selected dwellings were divided into 20 batches. The size of each batch ranged between 7,200 and 8,600 dwellings. Staggered start dates of about 3 to 4 days apart were planned for each of the 20 batches to spread out the workload for the three data collection modes.

As the pool of field interviewers tended to shrink towards the end of the fieldwork enumeration period, more dwellings were allocated in the earlier batches than in the last few batches. The allocation of more dwellings for enumeration in the first few batches helped field interviewers pick up the operational procedures and interviewing skills within a shorter period of time. Fewer cases towards the end also ensured that the remaining field interviewers were able to cope with the new batches and at the same time continue to revisit households that were not contactable from previous batches.

3.2 WORKFLOW OF CENSUS BATCHES

On average, three batches were rolled out within 2 weeks as illustrated in Figure 3.1. This staggered workflow was designed to optimise the use of resources. Apart from moderating the volume of Internet traffic to the Census Internet enumeration website and avoiding a sudden surge in usage should all respondents were to register at the same time, staggered batching also allowed for more effective case allocation and follow-up at the call centre and by field interviewers.

Figure 3.1: Flow of Records by Batch and Weeks after Census Launch

Batch	Weeks									
	1	2	3	4	5	6	7	…	n	→
1	Internet	Internet	Internet CATI	Internet CATI	Internet CATI	Field-work	Field-work	Field-work	Field-work	
2		Internet	Internet	Internet CATI	Internet CATI	Internet CATI	Field-work	Field-work	Field-work	Field-work
3			Internet	Internet	Internet CATI	Internet CATI	Internet CATI	Field-work	Field-work	Field-work
↓										
20									Internet	…

Respondents in each batch were first sent a notification letter informing them that they had been selected to participate in the Census 2010. They were given two weeks to provide their survey returns via the Internet. They also had the option to call the Census hotline to provide their information via a telephone interview. The Census hotline was manned by the CATI interviewers who would be able to conduct the Census survey over a telephone interview, answer queries about the survey and provide technical assistance for respondents using the Internet submission.

After the two-week period provided for Internet submission, the CATI system then began calling respondents who had not completed their returns. For respondents who had not completed their returns by the end of the second week, a reminder letter was sent. This reminder letter encouraged respondents to submit their survey returns via Internet or call the Census hotline for telephone interview. Respondents would still be able to submit their returns either through Internet or CATI during the third to fifth week.

A second reminder was sent at the end of the fourth week to alert respondents that survey officers could be arranged to visit them at their homes to assist them in completing their returns if they were not able to submit their returns via the Internet or telephone by the given deadline.

Records which were not completed via Internet or telephone interview by the end of the fifth week were assigned to survey officers for follow-up through face-to-face visits. Records were downloaded to the respective team members' Ultra-Mobile Personal Computers (UMPCs) to facilitate the conduct of field interviews. Field interviews took place from the sixth week. The Field team was given another four weeks to complete the survey returns in each batch.

All completed records flowed into the data processing system at the end of each day. This enabled the Data Processing team to start coding and editing the data as

and when the records were completed without having to wait for a specific batch to be completed and consolidated.

After stringent quality checks and the completion of data processing, multiple aggregated cross-tabulations were generated for analysis of data trends. Statistical reports on specific topics were prepared for public release.

3.3 INFRASTRUCTURE SET-UP FOR DATA COLLECTION

The Census 2010 data collection system was hosted at the Government Data Centre Shared Hosting Service (SHINE). SHINE is a one stop environment for government agencies to host and deploy their applications, e-Services and web sites. By riding on the centrally managed infrastructure for the public sector, lower operating costs were achieved with economies of scale. The system availability was also assured through a resilient central infrastructure. In addition, the round-the-clock operational support provided was essential to meet the needs of the Census Internet module which was available on a 24-by-7 basis.

Data collected from the Internet, CATI and Fieldwork were stored centrally for ease of data synchronisation. This was an enhancement from the last Census where two separate databases were used for the Internet/CATI and Fieldwork responses.

The design of a single database provided greater flexibility for households to switch their modes of submission during the survey period. It improved the overall productivity of the collection and provided convenience to households as there was no time lag in data update across the three modes of data collection. With more timely updates, communication with respondents also improved.

For the Department, there were also cost savings as the use of a single database reduced the costs in maintaining the multiple databases and servers and in transferring data across databases of different data collection modes.

To address the potential security concerns of having too much information in a single database, information was loaded in batches as and when each batch became available for enumeration. Upon receiving the complete information for a household, information for the household was moved from the data collection database to the internal data processing database to minimise the risk of exposure.

3.4 INTERNET DATA COLLECTION

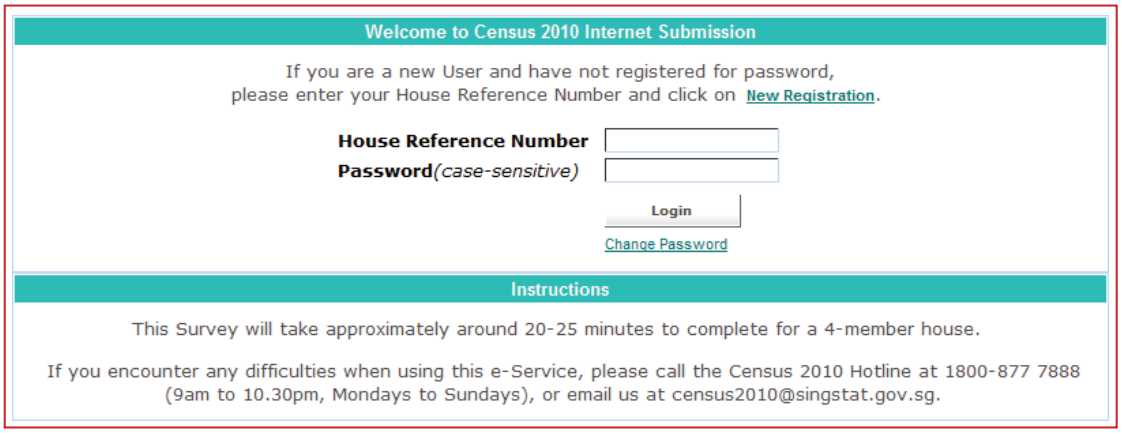
The system used for the Internet Data Collection was codenamed Electronic Submission Module (ESM). The design of the ESM built on the lessons learnt from the Internet submission systems used in Census 2000 as well as the General Household Survey 2005. Improvements were made to address feedback raised then and new features were introduced to meet the anticipated needs of the public.

3.4.1 Key Features

Self Registration and Creation/Reset of Password

To use the ESM for Census 2010, respondents first logged in using the House Reference Number (HRN) that was sent to them with the notification letter. One key feature of the ESM was the facility for respondents to register and create their own password (Figure 3.2). This reduced the risk of unauthorised access in the event that the notification letter was misplaced as the information in the letter itself would be insufficient to gain access to the household's Internet form.

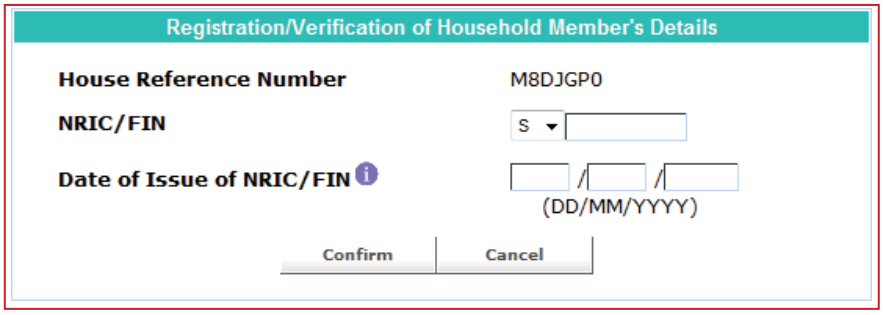
Figure 3.2: ESM Login screen



After entering the HRN and clicking on the "New Registration" link, the respondent would be prompted to provide the National Registration Identity Card (NRIC) or Foreign Identification Number (FIN) of any member living in the selected house and the corresponding Date of Issue (DOI) of that NRIC/FIN for verification (Figure 3.3).

In past surveys, respondents who were unable to provide the DOI of their NRIC/FIN such as those who had recently lost their NRIC/FIN or military personnel who might not have access to their civilian identity cards would not be able to register for the online Internet form. In Census 2010, supervisors were able to issue a Date of Registration (DOR) in place of the DOI to allow the respondent to proceed with the registration after an authentication process.

Figure 3.3: ESM Registration



Once the registration details were entered and verified, respondents would be prompted to create their own password. This password would be used for subsequent login. Respondents who forgot their password were able to change their password on their own by providing the original set of registration details. Allowing respondents to create and change their password minimised the turnaround time.

Respondents were able to start the Internet enumeration once the password had been created. If the NRIC/FIN and DOI entered earlier belonged to a registered person in the address, respondents would be able to retrieve an Internet form containing selected pre-loaded information about the household members. Otherwise, a blank form would be triggered.

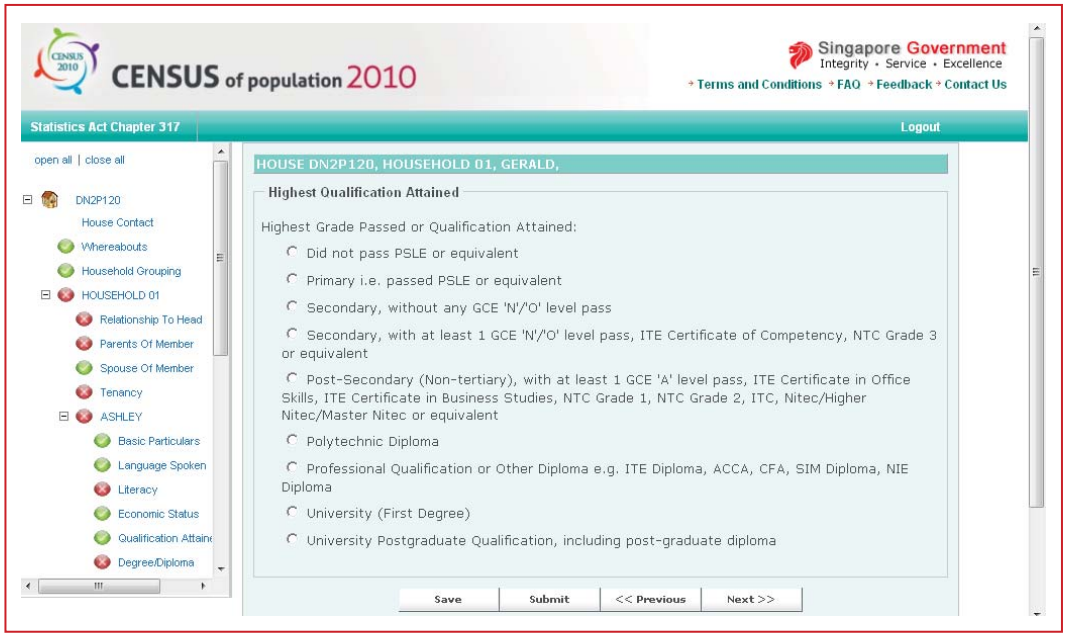
User-Friendly Form

The enumeration screens were designed to incorporate user-friendly features such as consistent placement of key functions, hyperlinks to "Frequently Asked Questions" and other relevant information such as "Glossary" which explained terms used in the survey form (Figure 3.4).

The left hand panel on the Internet screen enabled users to navigate to the intended screens directly, without having to use the "Next" or "Previous" buttons to move screen-by-screen. Screens that were completed were also marked with a tick in the left hand panel to allow respondents to monitor the progress and outstanding questions.

Automated branching of questions was in place to direct the respondent to the relevant questions. Respondents could also save their partially completed returns and re-visit them later to provide the remaining information.

Figure 3.4: ESM Enumeration screen with user-friendly features



Individual Member and Household Submission

Unlike in past surveys where all members in the same household had to submit their information in the same return, submission for individual member(s) in the household could be made in the Census 2010 Internet form. A respondent who completed the survey return for his/her part could submit his/her returns without the information for the rest of the household members being completed. Once his/her information was submitted, the rest of the members would not be able to view his/her returns. This enabled the respondent to keep his/her returns confidential from even those within the same household. Similar to individual member submission, separate submission at the household level could also be made if there were more than one household in the selected address.

Validation Prior to Submission

Upon submission, the system would trigger a series of validation checks for completeness and accuracy. Incomplete or erroneous fields were displayed as errors. Respondents were able to navigate to the missing/error fields directly to resolve them and submit successfully (Figure 3.5).

Figure 3.5: Screen showing submission status of survey returns

Short Message Service (SMS)/Email Reminders

Respondents could choose to receive SMS and/or email reminders before the due date for submission. This feature catered to respondents who were not able to complete their submission within one session. These SMS and email reminders were sent to the mobile numbers and email addresses provided by the respondents when they log on to complete the Internet survey forms.

Security

Data provided over the Internet were transferred via an encrypted and secured network. The IT infrastructure and system design were fully compliant with the Government IT security standards and best practices.

To prevent the possibility of online attack, additional security measures and procedures were built in. For example, the number of unsuccessful attempts for registration/login/change password was tracked. The house account would be locked if there were six unsuccessful login attempts. The CATI supervisors would only proceed to unlock the house account after completion of an authentication process.

A 20-minute session time-out feature was also built in as part of the security features. A prompt would appear after 15 minutes of inactivity to notify the respondent that the session was expiring. If there were no further activity for the next 5 minutes, the system would prompt "Session time-out" and log out automatically. This protected the data collected in the event that the respondent forgot to log out from the survey after use.

Hotline/Technical Assistance

To assist respondents who may encounter difficulties using the Internet form, the Census hotline was made available from 9am to 10.30pm daily, including weekends and public holidays. The operating hours were extended from 9pm during the conduct of Census 2000 to 10.30pm in Census 2010 to better support the respondents in their submission.

3.4.2 Additional Channels for Internet Submission

To expand the reach of the Internet submission and ensure that respondents were able to submit their Census returns via different channels, an Internet submission facility was set up at the Department of Statistics' office as well as at the 28 CitizenConnect Centres across Singapore.

Each CitizenConnect Centre is a one-stop centre for easy and convenient access to government services online. These centres open from 12pm to 8pm, Mondays to Sundays and are closed on public holidays. By leveraging on the CitizenConnect Centres, respondents who did not own a computer at home and those who were not so technically savvy were also able to submit their Census forms electronically.

3.4.3 Lesson Learnt

Online Registration Process

A number of respondents whose HRN ended with "0" mistook the number as the letter "O", causing problems in online registration. An additional message "*Note that the first five and last digits of the House Reference Number are numbers and the*

sixth character is an upper case alphabet. For e.g. 12345A6." was loaded in the login screen to highlight to the respondents on the format of the HRN. The use of ambiguous numbers/letters (e.g. "0" and "O") will be avoided for future surveys.

3.5 COMPUTER-ASSISTED TELEPHONE INTERVIEW

CATI is a tried-and-tested data collection method, having been successfully deployed thrice in the Census of Population 2000 and past General Household Surveys. Respondents, who were not able to complete their returns via the Internet, had the option to provide their returns over the phone with the assistance of telephone interviewers.

About 150 CATI interviewers were employed in 2 shifts per day, managed by a team of 15 Assistant Census Directors (ACDs) and Census supervisors.

3.5.1 Key Features

Important features and lessons learnt from previous deployment were integrated into the CATI system design for the Census 2010. Improvements were made to key features which included automatic branching of questions, online verification checks for completeness and automated dialling.

Automated Outbound Dialler

One key feature of the CATI system was the Automatic Outbound Dialler (AOD). The automated dialling system searched through telephone numbers from the list of available records and made calls based on a set of priority rules built into the system. When a respondent picked up the phone, the call was automatically routed to an available CATI interviewer who would interview the respondent. If the line was busy or there was no response, the system would search for the next available telephone number.

Switching between Inbound and Outbound Calls

The CATI system handled inbound and outbound calls, including appointment calls to households who had requested to be surveyed at a preferred date and time. As the pool of CATI interviewers were trained to handle both inbound and outbound calls, manpower resources could be optimised with the switching of roles to meet the demand of the day. For example, if the inbound call traffic was slow during the inbound off-peak period, the automated dialler would identify available CATI interviewers and assign them with outbound calls. Therefore, the same interviewer could be involved in a variety of tasks – handling hotline enquiries which may include basic technical support to Internet users, attending to inbound calls for enumeration, managing outbound calls, as well as providing assistance to field interviewers who called the hotline for updates to their assigned records.

Management of Call Load

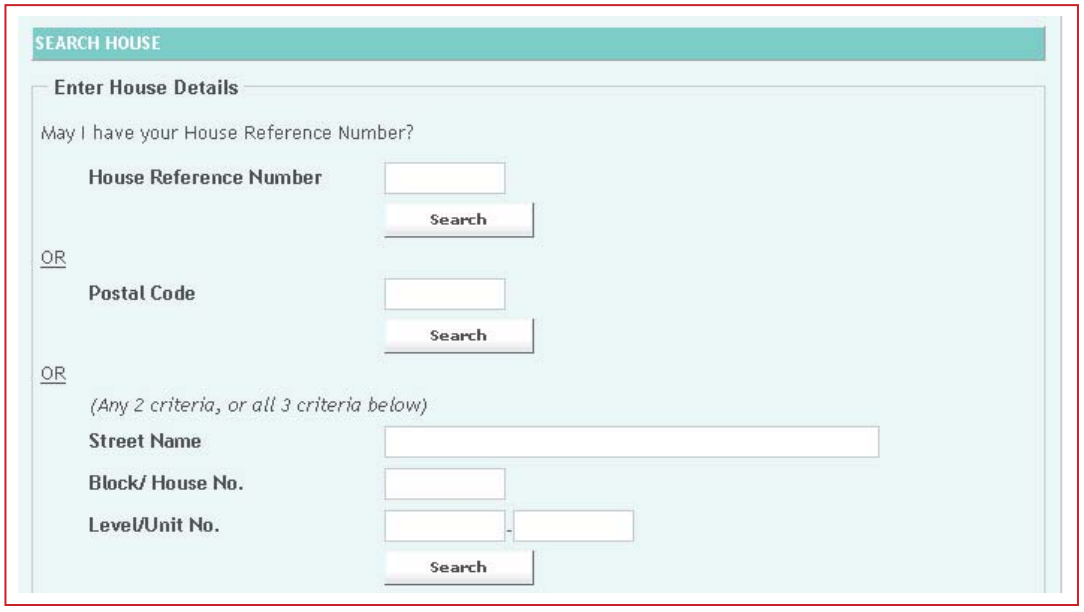
Inbound call traffic and patterns were monitored closely through the generation of management reports from the system during CATI operating hours. Availability of real-time reports enabled CATI supervisors to adjust the system settings to appropriately handle the call volume accordingly. For peak periods with high volume of inbound calls, more CATI interviewers were allocated to receive inbound calls to ensure that respondents calling in were attended to.

For inbound calls received before or after the CATI operating hours, or during peak periods when there were no available CATI interviewers to answer their calls, the respondents could leave their telephone number. The CATI system would automatically schedule to call them at the start of the next working day or as soon as a CATI interviewer was available.

Authentication Process

There were slight variations between handling an inbound call and an outbound call. For an inbound call, it was necessary for the CATI interviewer to first verify the selected address. The following screen (Figure 3.6) would be displayed for inbound calls:

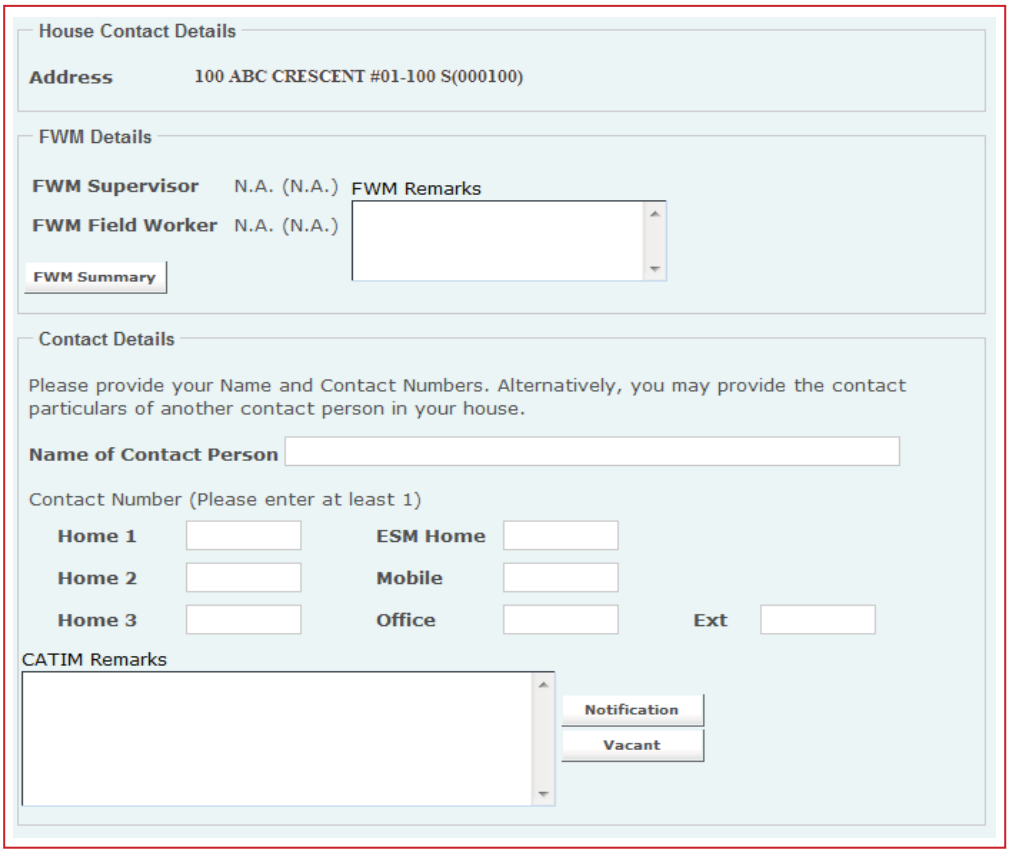
Figure 3.6: CATI Search screen for inbound calls



The CATI interviewer would ask for the HRN which was printed on the notification letter sent to the household as part of the notification package. As it was common for respondents to misplace or forget the HRN, an alternative verification mechanism was put in place to allow the CATI interviewer to search for the address of the caller using the 6-digit Postal Code or a combination of Street Name, Block/House number and Level/Unit number. After the address was verified, the following screen

(Figure 3.7) would be displayed. The CATI interviewer would proceed with the enumeration.

Figure 3.7: CATI Enumeration screen on contact details



For an outbound call, the address corresponding to the telephone number would already be known since the call was made by the system. The CATI interviewer would quote the HRN and confirm the address with the respondent before proceeding to conduct the interview.

CATI interviewers were adequately trained to multi-task and perform duties ranging from conducting interviews, answering queries, to providing support to the field team. With the assistance of the CATI interviewers, the survey could be completed and submitted within 20 minutes for a typical 4-member household. CATI also proved to be a useful mode of survey for respondents who were illiterate or have no Internet access.

Supervisory Support

Supervisors were available during the operations to guide interviewers and handle more complex queries. An administrative summary screen (Figure 3.8) enabled supervisors to access various details on the house record (e.g. first registrant's information, Internet due date, survey mode and whether house account had been locked etc.). Such information enabled the supervisors to assist the respondents to resolve the problems they encountered.

Figure 3.8: CATI Administrative Support screen

Extended Operating Hours

Experience in previous surveys revealed an emerging trend of respondents returning home at later hours. To cater to this group of respondents who reached home in the late evenings, the operating hours of the Census 2010 call centre were extended from 9pm in Census 2000 to 10.30pm. Respondents could contact the hotline from 9am to 10.30pm daily, including weekends and public holidays. The extension of operating hours was in line with the Department's commitment to provide better service and support to our survey respondents.

User-Friendly Form

The question flow in the survey form used at CATI was the same as the Internet form. With automatic branching of questions, the system displayed relevant questions pertaining to the household and its members based on their profile. The online verification checks for completeness highlighted missing fields or data discrepancies to the CATI interviewer, so that clarifications could be made with the respondent before the survey was submitted.

The consistency in interface with the Internet form also facilitated the CATI interviewers' support to the Internet respondents. The CATI interviewers were able to provide screen-by-screen guidance that corresponded to what the respondents were seeing.

Co-ordination with Fieldwork and other Operations teams

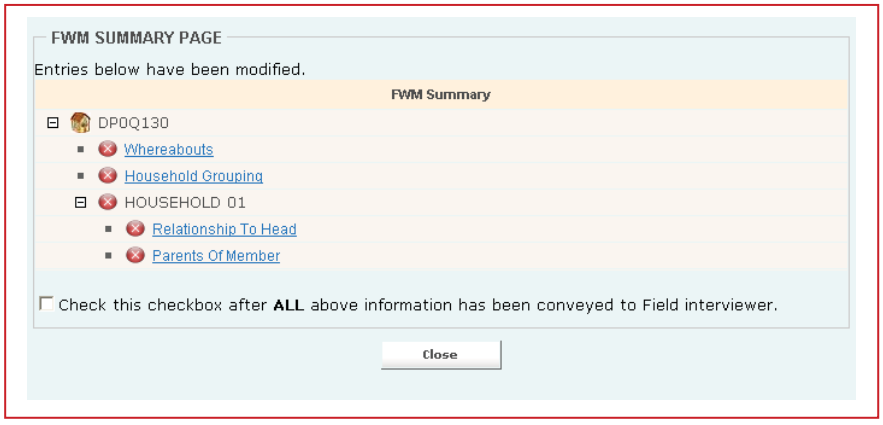
As respondents could still call in to the hotline to furnish information or submit their returns after the field interviews started, it was important to ensure that timely

updates were provided to the field team. SMS alerts were sent to update field interviewers on the completion status of a house record which had been allocated to them. Field supervisors were also notified through SMS of the updates to assist them in monitoring the interviewers' duties and work progress.

When alerted that new information had been furnished via CATI, the field interviewer would contact the Census 2010 call centre to obtain the latest updates before visiting the respondent at his/her house. A Fieldwork Summary screen at CATI allowed CATI interviewers to access the screens which have been updated over CATI (Figure 3.9). The CATI interviewers would convey the updates displayed in the screen to the field interviewers. This automated the process of alerting field interviewers and allowed them to obtain the latest information pertaining to the household. Updates were provided on a regular and frequent basis, in intervals of fifteen minutes. This ensured minimum time lapse for the updates.

Besides working closely with the field team, the CATI team also worked with the Data Processing team on clarifications and amendments to previously submitted records. Feedback was also sought and provided to the Quality Assurance team to improve the quality of the data collected and effectiveness of the training on an on-going basis.

Figure 3.9: Fieldwork Summary screen



3.5.2 Lessons Learnt

There were several critical lessons that could be learnt from conducting Census 2010 using CATI:

Design of House Reference Number

A handful of HRNs were very similar (e.g. "12345N3" and "12345M3", "67890F1" and "67890S1"). Some CATI interviewers encountered difficulty differentiating the sound of certain letters in the HRNs reported by the respondents over the telephone. For future surveys, the use of such letters in HRNs could be avoided.

Fast Downloading of Screens

CATI interviewer often had to read out the questions on the screen promptly during the interview with the respondent. Hence if the system response time was slow, the survey time would be prolonged and respondents would get impatient. To prevent this, it was important to stress-test the system to ensure that there was no lag in response time even during peak enumeration periods.

Role of the Logistics team

As there was much co-ordination required across the various teams collecting and processing the Census data, the Logistics team played a pivotal role in supporting the CATI and Fieldwork teams. Besides overseeing the recruitment, they kept track of the notification letters bounced back from delivery for further follow-up at CATI, handled the requests for letters to be resent and managed the respondents who walked in for enumeration. The set-up of the separate support team contributed to the smooth progress of the collection and processing teams which could then focus on the respective tasks.

3.6 FIELDWORK OPERATIONS

To facilitate fieldwork operations, selected dwellings for Census 2010 were grouped into 5 geographical divisions, namely: North, North-East, East, West and Central. Each division was headed by an ACD. Within each division, there were 4 sub-divisions, each managed by a field supervisor.

Field operations were carried out from 10 regional offices (ROs) with about 140 field interviewers, managed by a team of 20 field supervisors.

3.6.1 Key Features

Use of Ultra-Mobile Personal Computers

One of the major improvements in Census 2010 fieldwork operations was the deployment of UMPCs. A fieldwork application was developed for field interviewers to carry out their enumeration. Unlike the traditional paper and pen approach, the use of UMPCs provided logistic convenience, eliminating the need for printing and transportation of hardcopy survey forms.

Automatic branching of questions directing the field interviewers to only the required questions sped up the enumeration process. Completeness and consistency checks built into the application reduced accidental omissions of questions and allowed clarifications to be made on the spot. Back-end processing efforts were also reduced with electronically coded data captured from the enumeration. Furthermore, information captured was kept within the secure handling of the UMPC as data enumerated were encrypted within the application.

Fieldwork Management System

A computerised Fieldwork Management (FWM) system was also developed for supervisors to allocate records, transfer records and keep track of the field interviewers' workload and the status of the records assigned to them. In addition, field supervisors were able to view records that would potentially flow to fieldwork one week prior to the start of fieldwork phase for planning purposes.

The field interviewers met their supervisors at their respective ROs twice a week for progress updates and to synchronise the data collected in the field into the Data Collection Database. The ACDs and supervisors provided close supervision and guidance to the field interviewers to ensure that the field operations were carried out properly and expeditiously.

To further facilitate data collection at field, supervisors carried out field reconnaissance to familiarise themselves with their fieldwork areas and make necessary arrangements for the field interviewers. For example, arrangements for access into private condominiums were made with the respective condominium management committee prior to the field interviewers' visits where necessary.

3.6.2 Fieldwork for Large Households

The Survey of Large Households (SLHH) was conducted prior to the main Census 2010 to improve the efficiency during the fieldwork operations. The SLHH covered households in the Census 2010 sample that were large in size as these houses were potentially more difficult to enumerate. The large size of such households also meant that the time taken for them to complete their survey via the Internet and CATI would be significantly higher than the average households. Additional resources would also be required to cater to them if they were to be included in the main survey.

The SLHH was conducted through face-to-face interviews over a period of one week, from 22 to 26 February 2010. All dwelling units containing 15 or more registered individuals in the Census 2010 selected sample were identified.

This survey also provided practical training for both the field and CATI supervisors. The field supervisors gained first-hand experience of the field conditions in advance of the Census and the CATI supervisors were able to gain experience in managing respondents' feedback and queries over the telephone.

3.6.3 Lessons Learnt

Like the ESM and CATI, there were several critical lessons that were learnt from Fieldwork:

Deployment of UMPC for Enumeration

The use of UMPCs greatly improved efficiency in fieldwork operations. Supervisors no longer had to print out the available information onto Household and Individual forms for field interviewers to conduct interviews. It was not necessary to transport the forms back to Census Headquarters via courier service and to scan the forms electronically. For the field interviewers, the burden of carrying hardcopy forms around and ensuring the completeness of all the relevant fields on the hardcopy forms were relieved from them with the use of UMPCs.

Improved FWM System for Fieldwork Supervisors

Using the FWM system, supervisors were able to view pre-allocated house records one week before actual allocation to the interviewers and this enabled them to carry out effective planning of records. The FWM system also allowed supervisors to view the list of outstanding and submitted records assigned to the interviewers which was helpful in monitoring the fieldwork progress and rendering timely assistance to the interviewers whenever necessary.

Knowledgeable and Confident Field Interviewers

As fieldwork was the most difficult phase of the data collection, it was important for field interviewers to be knowledgeable and confident to secure the respondents' co-operation. A rigorous 3-day training program was organised to equip field interviewers with the essential understanding and skills before being deployed to the field. Upon actual deployment, the respective field supervisors and interviewers met frequently to discuss problems faced during the survey and possible solutions at their ROs. These enhanced the knowledge and confidence of field interviewers for Census 2010.

Availability of Census 2010 Publicity Pamphlets for Interviewers

The availability of the Census 2010 publicity pamphlets in 4 different languages (English, Chinese, Malay and Tamil) enabled field interviewers to explain clearly to the respondents about the conduct of Census 2010. The publicity pamphlets coupled with the publicity posters at bus stops and Mass Rapid Transit (MRT) trains as well as advertisements on television and radio created greater awareness of the survey to the respondents.

3.7 DATA PROCESSING

The Census 2010 data processing systems incorporated enhanced features to streamline operations, such as batch coding, batch validation and sorting of records for coding, to reap productivity gains.

Census returns submitted via the Internet, CATI and field enumeration were captured electronically and updated into the Census database directly without the need for separate data entry back in office. Trained coders performed online computer-assisted coding of data on occupation, industry and education while editors checked on the consistency and completeness of survey returns. The final processing of the data involved the editing and verification of records to ensure consistency and good quality of the data.

About 100 coders and editors were employed in 2 shifts per day, managed by a team of 8 ACD and Census supervisors.

3.7.1 Coding of Vocational Qualification, Industry and Occupation

For the main Census 2010, the data collected were captured in pre-coded categories except for selected descriptive fields which included vocational qualification, industry and occupation. These fields were reported by respondents in text format and needed to be coded in the office.

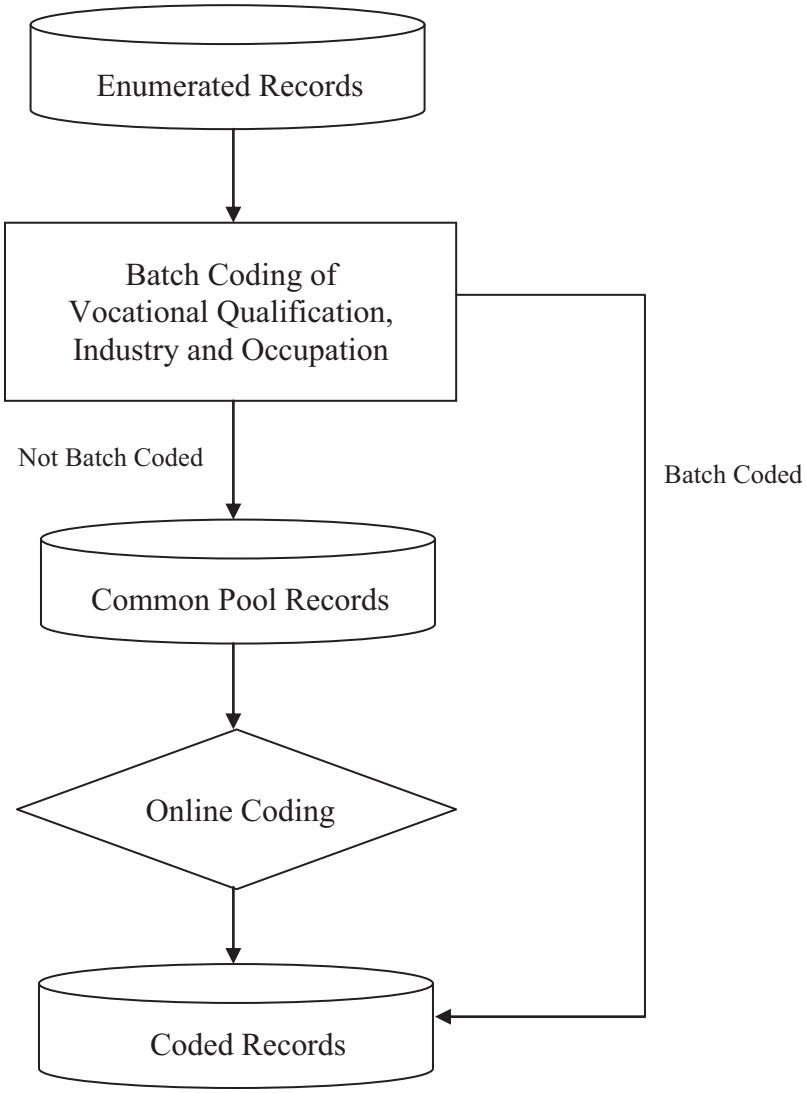
Coding was the first phase in data processing. This involved the coding of descriptive information collected on vocational qualification, industry and occupation. To expedite the process, coding of vocational qualification, industry and occupation was further organised into batch coding and online coding (Figure 3.10).

Batch coding was an automated process during which appropriate codes were assigned by the application program based on the descriptive information captured on vocational qualification, industry and occupation and in-built coding rules. Records that were distinct and non-ambiguous were coded automatically. Batch coding of vocational qualification was performed first, followed by industry and then occupation.

To facilitate batch coding, input files containing descriptive information on vocational qualifications, industries and occupations with their corresponding codes were prepared. The files were collated using descriptions collected in past Censuses and household surveys as well as the latest occupations and industries listed in the Singapore Standard Industrial Classification (SSIC) and the Singapore Standard Occupational Classification (SSOC).

During the batch coding process, the descriptive information captured for the Census 2010 was matched against the input files. If a complete match was found, the record was assigned with the specific code. If no match was found during batch coding, the record flowed into a “Common Pool” for online coding. About 18 per cent of records which required coding were assigned vocational qualification/industrial/occupational codes through batch coding. This was a marked improvement from the corresponding rate of 6 to 7 per cent during the Census 2000.

Figure 3.10: Coding Workflow



Sorting and batch allocation was performed for online coding to improve the productivity of the coding staff. Records with similar characteristics were grouped and allocated to the same coding staff where applicable. For online vocational qualification coding, the records were first sorted by Batch Number and followed by Highest Academic Qualification Attained (HAQA). For online occupation coding, the records were sorted by Batch Number and Employment Status followed by broad Industry Group or HAQA, depending on the Employment Status.

During the online coding process, coding staff used a search facility to retrieve relevant information from the input files for determination of the appropriate codes. This search facility used the "key word" search method to find an appropriate match. The results of the search were displayed on-screen for the staff to assess and assign the most appropriate code. For example, a staff could search using a key word to

retrieve all commonly used occupation titles that contained this key word and the corresponding SSOC codes.

After coding was completed, the records flowed to the editing process which was the second phase of data processing.

3.7.2 Data Verification and Editing

Data verification and editing were undertaken to ensure all records were accurate before they were used for compilation of aggregated statistics. The main processes (see workflow Figure 3.11) were as follows:

- Batch Validation
- Online Verification and Editing
- Duplicate Checking

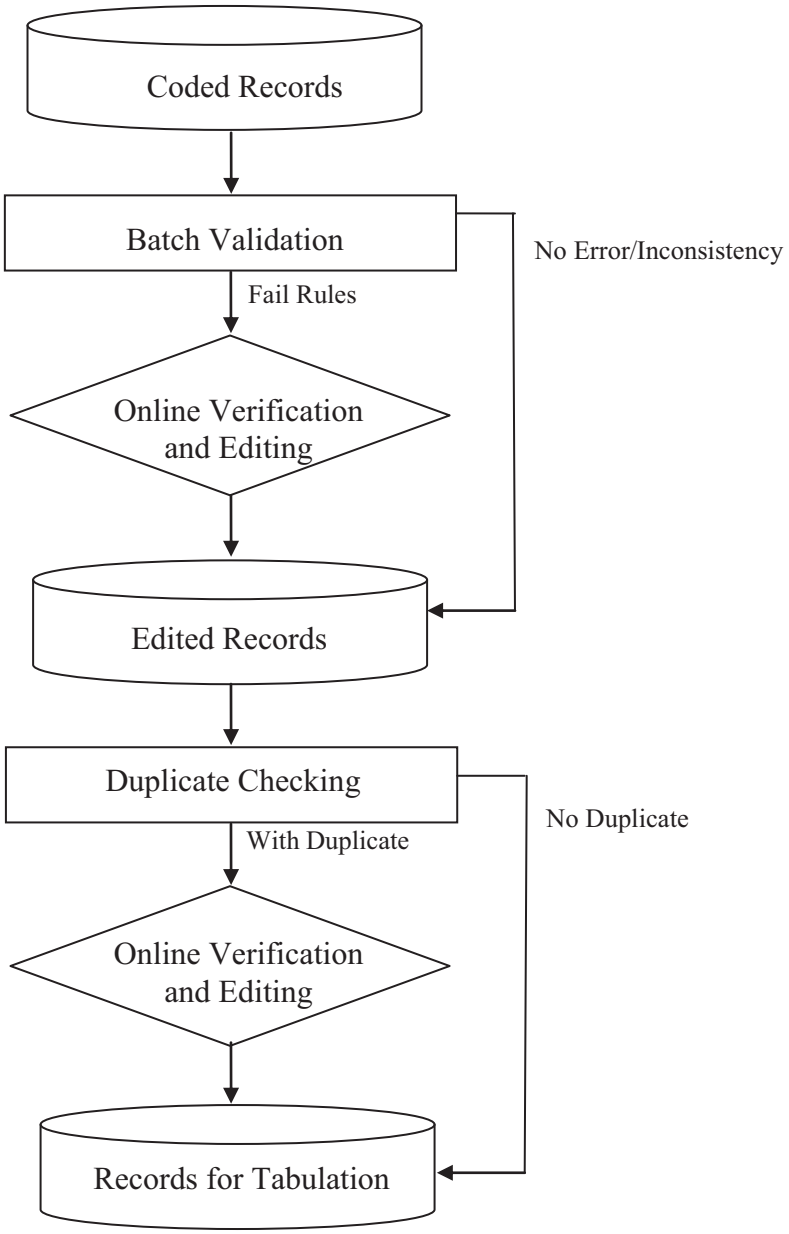
Batch validation was an automated process during which all records went through a series of stringent checks by the computer. The "error" and "consistency" rules were consolidated from past experiences with censuses and household surveys. "Error" rules checked primarily for missing key information which should not be left blank and validity of codes. They were also designed to detect records with entries for two or more data items which were logically impossible. "Consistency" rules checked for scenarios which were unlikely to occur but could still exist. Examples of such rules include small age difference between parent and children, and attainment of university qualifications at relatively young age. Records which failed error or consistency rules during batch validation were flagged out and retrieved by editing staff for online verification and editing.

During the online verification and editing process, editing staff scrutinised and corrected the errors and inconsistencies that were highlighted for each record on-screen. Where necessary, editing staff would contact respondents to seek clarifications. For records which failed rules pertaining to vocational qualification, industry or occupation codes, editing staff could access the relevant functions in the online coding systems to help them to review and amend the codes, if required. The editing staff could iteratively check their records against the error and inconsistency rules after making amendments to them, and this checking process would continue until the records were error-free and all inconsistencies, if any, had been verified.

After all the records were edited, a duplicate check was conducted to retrieve records of individuals who were enumerated more than once during data collection. For example, this could happen when an individual who had been enumerated in one house was enumerated again after he shifted to another house within the survey period. For such cases, the additional records were deleted to prevent double

counting. The retained records were then passed through the online verification and editing for another round of verification of errors and inconsistencies.

Figure 3.11: Editing Workflow



3.7.3 Quality Assurance

As part of data quality assurance, a number of cross-variables consistency checks were performed for related items of the edited records to identify possible misreporting. Examples of such checks include persons who reported multiple modes of transport with short travelling time or elderly persons who indicated main source of financial support as employment/business income but reported no income from work.

Records which failed the quality assurance checks flowed back to the editing team for further verification and editing. Once all the records were verified to be error-free, they were used for data compilation and analyses. The quality assurance checks served to strengthen the data verification process and enhance the quality of data captured.

3.8 RESPONSE RATE AND MODE OF RESPONSE

The overall response rate for the Census 2010 was about 98 per cent. Households that were non-contactable during the Census survey period and those who provided partial information but did not complete the survey were included as non-responding units.

Households who responded to the Census through the CATI formed the largest group at 46 per cent of responded households. Another 38 per cent submitted their Census returns via the Internet while the rest (16 per cent) were surveyed through face-to-face interviews. More details on the take-up by the different modes of data collection are discussed in Appendix N.

3.9 DATA DISSEMINATION

The register-based Census coupled with the tri-modal data collection strategy and expeditious completion of data processing contributed towards timely dissemination of the Census 2010 results.

3.9.1 Data Tabulation System

An in-house project was undertaken to develop a reliable and easy-to-use tabulation system for the Census 2010. The tabulation system uses Microsoft Excel macro software as an interface to compile the aggregated data at SAS server before exporting to Excel spreadsheet. As this system was riding on SAS engine, it could process records and perform tabulations swiftly.

The tabulation system has the following advantages compared with software developed in earlier major projects:

a) The system could generate different types of statistical outputs, through built-in templates (or options) to cater to different data requirements. The statistical outputs include cross-tabulations on frequency count, percentage distributions and summary statistics (e.g. mean, median). A maximum of 4 variables could be specified to produce the required report.

b) SAS output could be converted directly to Excel format. This was faster than if SAS/E-Guide were to be used and for data to be manually exported to Excel for tabulation. The final output could be used to do additional grouping or calculation for analysis.

c) The incorporation of user-friendly templates in the system enabled statistical tables to be produced in a ready-to-publish Excel format and reduced the need for further formatting.

3.9.2 Statistical Releases

From the register-based Census, the Department published the report "Advance Census Release" in end-August 2010 on key trends in the population size, growth, and broad geographic distribution of the Singapore population as at Census Day, 30 June 2010. This was followed by a Statistical Newsletter article in September 2010 providing more register-based information by geographical breakdown (Appendix O).

Following the completion of data processing for the Census sample enumeration, the Department released a series of topical Statistical Releases (SR) on detailed Census results during the period of January-February 2011, as follows:

Publication	Topic	Release Date
SR No. 1	Demographic Characteristics, Education, Language and Religion	12 Jan 2011
SR No. 2	Households and Housing	15 Feb 2011
SR No. 3	Geographic Distribution and Transport	24 Feb 2011

These Statistical Releases present the broad trends and changes between 2000 and 2010. The reports also contain key indicators, comprehensive and detailed tables, charts and descriptions of the concepts and definitions which serve to meet the needs of the general public for a wide range of data on Singapore's population and households. They serve as comprehensive reference sources for planners, researchers and other data users requiring in-depth data for their analyses.

The Census 2010 reports are available for free download from the release date, on the Department's website at www.singstat.gov.sg. The printed versions are available for sale.