

ITEM FOR FINANCE COMMITTEE

**CAPITAL WORKS RESERVE FUND
HEAD 710 - COMPUTERISATION
Census and Statistics Department
New Subhead “Computer Equipment and Services for the 2011 Population
Census”**

Members are invited to approve a new commitment of \$85,273,000 for acquiring computer equipment and services for the 2011 Population Census.

PROBLEM

Given the enormity of the operation and the huge volume of data to be processed, it would not be practically possible to conduct the 2011 Population Census (11C) without adequate computer support.

PROPOSAL

2. The Commissioner for Census and Statistics, with the support of the Secretary for Financial Services and the Treasury and the Government Chief Information Officer, proposes to acquire computer equipment and services for developing an integrated computer system to support various activities in the 11C.

JUSTIFICATION

Computer Support for the 11C

3. The 11C is a large-scale and complex operation involving enumeration of some 2.4 million households during the 45-day data collection period from 19 June to 2 August 2011; processing of completed questionnaires in

/around

around six months; and dissemination of the 11C results in the form of statistical reports from February 2012 to March 2013. As borne by operational experience in past censuses, computer support is essential to ensure the smooth implementation of a census of such a scale.

4. In conducting the 11C, Census and Statistics Department (C&SD) will adopt a Long Form questionnaire on detailed characteristics for a sample of 10% of the entire population and a Short Form questionnaire on basic characteristics for the remaining population. Different from previous population censuses in which data were collected basically by face-to-face interviews, a new multi-modal data collection approach will be introduced in the 11C. Under the new approach, C&SD will, during the first stage of the data collection period (19 June to 15 July 2011), encourage self-enumeration by requesting participating households to return their completed questionnaires by post or via the Internet, while the traditional interviewer method will be activated for enumerating the remaining households in the second stage (16 July to 2 August 2011). An integrated computer system is therefore needed to consolidate the returns collected from different modes under the new multi-modal data collection approach.

5. In 2008, C&SD commissioned a feasibility study to examine the technical requirements and financial implications of the proposed computer system. A Business Process Re-engineering (BPR) study was also conducted to identify opportunities for streamlining and improving the workflow of the 11C.

The Proposed Computer System for the 11C

6. The feasibility cum BPR study, completed in February 2009, proposes to “recycle” for the 11C the computer system developed for the 2006 Population By-census (06BC) to the extent possible. At the same time, the study recognises that redevelopment/upgrading of the 06BC systems will be required to meet the new requirements of the 11C (in particular the adoption of the new multi-modal data collection approach and the need to survey the entire population in the 11C) and to provide increased functionalities to ensure smooth operation of the 11C (in terms of better data quality and security control, more effective handling of public enquiries and fieldwork management as well as integration among various sub-systems). The study therefore recommends that a new 11C computer system should be developed using a combination approach involving enhancement of two sub-systems of the 06BC, re-development of its another 14 sub-systems, and development of two new sub-systems. This combination approach represents the most cost-effective solution in developing a computer system to support the 11C.

New Functions of the 11C Computer System

7. To provide better services to the public and to improve the efficiency of the 11C operation, the proposed computer system will be equipped with the following new functions/features –

(a) *Improving data accuracy*

In order to improve data accuracy, data obtained from various channels will be checked and validated on a daily basis^{Note} so that erroneous entries can be clarified and rectified with the respondents within the census period as far as possible.

(b) *Improving efficiency of temporary staff recruitment and training processes*

To support the field work operation of the 11C, C&SD needs to recruit and provide training for some 20 000 temporary staff. To streamline the recruitment and training processes and achieve greater cost effectiveness, C&SD will establish an on-line e-Recruitment module on the Internet to replace traditional paper applications. The on-line system will accept applications, release appointment results and arrange training schedules for successful applicants electronically. With the e-Recruitment module, we expect that the unit recruitment cost will be reduced by some 20%.

(c) *Improving field operational efficiency*

Intelligent Character Recognition technology will be extended to capture enumeration progress for better fieldwork monitoring. This will help avoid transcription errors and repeated checking of the progress updates, thereby shortening the processing time for generating progress updates. With more extensive use of information and communication technology in fieldwork management (such as providing more timely progress updates within one day from the fieldwork), the field operation can be managed more efficiently and the enumeration progress monitored more closely.

(d) *Launching the e-Reporting mechanism*

As mentioned in paragraph 4 above, on-line e-Reporting will be made available to all households in the first stage of the data collection period. Respondents can make use of the on-line system

/to

^{Note} In the past, computer-based data validation was done after the census period.

to access and complete the e-questionnaires in private and at their own pace. This can also save the interviewer manpower in the second stage. We expect that the take-up rates will be about 10% and 5% for Short Form and Long Form questionnaires respectively.

(e) ***Improving incident handling and reporting***

To improve our performance and responsiveness in incident handling, C&SD will set up workflow functions and establish a repository comprising Q&As and guidelines to enhance reporting efficiency and support more effective incident handling. The repository can facilitate staff members to quickly retrieve the records of past cases and take them into account in considering how best to handle evolving cases.

(f) ***Enhancing the quality assurance mechanism***

Apart from conducting quality checks on selected cases as in previous censuses, C&SD will further enhance the quality assurance mechanism to enable prompt identification of dubious cases, as well as timely review of “non-contact” cases and analysis of checking results in order to facilitate follow-up actions to be taken expeditiously and reduce the number of “non-contact” cases effectively.

(g) ***Strengthening the questionnaire tracking mechanism***

Under the multi-modal data collection approach, C&SD will establish a central system to record the whereabouts of completed questionnaires across different operations. This will help track the status of returned questionnaires and maintain a systematic flow of information from one system to another to avoid any mis-placement of questionnaires during transit.

Cost and Benefit Analysis

8. As mentioned in paragraph 3 above, given the large scale and complex operation of the 11C, it would not be practically possible to conduct the census without efficient computer support. The proposed system will be critical for the smooth conduct of the 11C and thus the production of good quality statistics for use by a large number of users in both the public and private sectors.

9. The proposed 11C computer system is designed to meet all the operational, timing and data quality requirements of the 11C project. It is designed using a combination approach as explained in paragraph 6 to maximise cost effectiveness. Besides, with the implementation of the new multi-modal data collection approach which would be made possible only by the proposed 11C computer system, the requirement for temporary field workers can be reduced by 3 000 with cost avoidance of \$27 million in the 11C when compared with the 2001 Population Census (01C), even though the number of households to be enumerated will be increased from 2.053 million in 01C to around 2.376 million in the 11C. As set out in paragraph 7(b) above, the proposed system will also help to reduce the unit cost of recruiting temporary staff.

Future Use of the 11C Computer System

10. The various components of the proposed computer system will be retained for use after the 11C operation. They will be used to support the 2016 Population By-census (16BC) as long as there are no fundamental changes in the business requirements and the performance of the 11C system can satisfactorily meet the operational requirements of the 16BC.

FINANCIAL IMPLICATIONS

Non-recurrent cost

11. We estimate that implementation of the proposal will incur a non-recurrent cost of \$85.273 million over a five-year period from 2009-10 to 2013-14 for the acquisition of computer hardware, software and related services. Detailed breakdown is as follows –

	2009-10	2010-11	2011-12	2012-13	2013-14	Total
	\$'000	\$'000	\$'000	\$'000	\$'000	\$'000
(a) Hardware	5,697	5,906	743	743	743	13,832
(b) Software	804	7,705	865	865	865	11,104
(c) Site preparation	2,500	500	-	-	-	3,000

/(d)

	2009-10 \$'000	2010-11 \$'000	2011-12 \$'000	2012-13 \$'000	2013-14 \$'000	Total \$'000
(d) Implementation services	500	7,057	14,759	500	-	22,816
(e) Contract staff	3,195	9,655	10,768	1,307	495	25,420
(f) Miscellaneous expenses	341	820	2,089	160	112	3,522
(g) Contingency	913	2,215	2,046	250	155	5,579
Total	13,950	33,858	31,270	3,825	2,370	85,273

12. As regards paragraph 11(a) and 11(b), the costs are for the acquisition of computer hardware and software. The hardware will include 20 computer servers, about 700 personal computer workstations, peripherals, network equipment and data communication facilities. The software will include operating systems, database management systems, application development tools, e-Reporting software and incident handling software, etc.

13. As regards paragraph 11(c), the site preparation cost is for the setting up of a server room and network connection in the 11C Office, the Central Processing Area and the Packing Site.

14. As regards paragraph 11(d), the cost is for hiring of services for equipment installation, system configuration, system development and customisation, data capturing services for completed questionnaires, and server hosting services.

15. As regards paragraph 11(e), the cost is for hiring contract staff services for carrying out the system implementation work, including defining user requirements, developing the corresponding sub-systems, providing support during production run, and monitoring the performance of the contractor for implementation services.

16. As regards paragraph 11(f), the cost is for rental of data communication link, acquisition of consumables and staff training.

17. As regards paragraph 11(g), the cost represents a 7% contingency of the items set out in paragraph 11(a) to (f).

/Other

Other non-recurrent cost

18. The implementation of the project will entail non-recurrent staff cost of \$5.066 million for the creation of three time-limited civil service Analyst/Programmer (AP) posts from April 2010 to August 2012. These civil service AP officers will be responsible for project planning and monitoring, system design, procurement and implementation, as well as liaison with users, other government departments and vendors.

19. The above non-recurrent staff cost will be absorbed by C&SD. The cost breakdown by years is as follows, –

	2009-10 \$'000	2010-11 \$'000	2011-12 \$'000	2012-13 \$'000	2013-14 \$'000	Total \$'000
Staff Cost	-	2,096	2,096	874	-	5,066
Total	-	2,096	2,096	874	-	5,066

Recurrent cost

20. The cost required for maintaining the computer equipment during the five-year project period has been included in the non-recurrent cost of the project. The recurrent maintenance cost beyond the project period will be absorbed by C&SD.

IMPLEMENTATION PLAN

21. Subject to approval of funding, we plan to adopt the following implementation schedule –

Major Activities	Target Completion Date
(a) Tendering and Procurement	March 2010
(b) System Design and Development	March 2011
(c) Trial Run and Fine-tuning	December 2010
(d) User Acceptance Test	May 2011
(e) Census Field Operation	August 2011
(f) Data Processing and Dissemination of Summary Results	February 2012
(g) Production and Dissemination of Other Census Results	March 2013

/PUBLIC

PUBLIC CONSULTATION

Encl. 22. On 5 January 2009 we briefed the Legislative Council Panel on Financial Affairs on our planning work for the proposed 11C. For Members' information, we set out at Enclosure supplementary information in response to questions raised by Panel Members on the proposed 11C. On 4 May 2009 we consulted the Panel on the proposed 11C computer system. Members generally supported the proposal.

BACKGROUND

23. Population censuses were conducted in Hong Kong in ten-year intervals in 1961, 1971, 1981, 1991 and 2001 and population by-censuses were conducted in the middle of the intercensal period in 1966, 1976, 1986, 1996 and 2006. The next population census will be conducted in 2011.

24. The aim of conducting population censuses/by-censuses is to obtain up-to-date benchmark information on the socio-economic characteristics and geographical distribution of the population. They provide benchmark data for studying the direction and trend of population changes. The data are key inputs for making projections concerning population, household, labour force and employment. Population censuses/by-censuses differ from other general household sample surveys in their sizable scale, which makes it feasible to provide statistics of high precision even for population sub-groups and small geographical areas. Such information is vital to the Government for planning and policy formulation, and important to the private sector and academia for business and research purposes.

25. On 25 June 2004, FC approved \$32,314,000 (*vide FCR(2004-05)19*) for acquiring computer equipment and services for the 06BC.

**Supplementary Information
in Relation to the Conduct of 2011 Population Census**

(a) Proposed adoption of a sampling fraction of one-tenth for the Long Form in the 2011 Population Census (11C)

According to the experience of the 2006 Population By-census in Hong Kong, a sampling fraction of one-tenth of the population is able to provide the required precision for supporting relevant statistical inference and analysis. The sampling fraction adopted by other jurisdictions varies depending on the extent of variation in population characteristics and the level of details required for data analysis. Same as Hong Kong, Korea has adopted a sampling fraction of one-tenth for its Long Form.

(b) Means of data collection

2. Electronic means of data collection would be adopted as part of the multi-modal data collection approach in the 11C. Supporting computer systems and applications would need to be set up to facilitate completion of questionnaires, particularly in respect of the more complicated questions, and their return to the Census and Statistics Department (C&SD) via the internet. In general, electronic means is considered more suitable for the Short Form which involves a few simple questions on basic demographic characteristics. As for the Long Form which requires provision of more detailed information from the respondents, it is anticipated that the internet submission rate may be lower. Experience in other developed countries like Canada indicates that about 10% to 20% of the households would complete self-enumeration through electronic means.

3. Telephone interview is not considered a suitable means of data collection for 11C, due to the number and complexity of questions in the Long Form which would make it difficult for enumerators to complete the interview over the telephone. Besides, C&SD does not have ready information of telephone numbers of all households.

(c) Information on the underprivileged such as people living in poverty, new arrivals, ethnic minorities, sexual minorities and people with disabilities

4. In the past population censuses/by-censuses, C&SD has prepared thematic reports on the socio-economic characteristics of different population sub-groups. For example, thematic reports on persons from the Mainland having resided in Hong Kong for less than seven years, single parents, ethnic minorities as well as thematic reports on household income distribution were published after the 2006 Population By-census.

5. In line with the international practice, information on sensitive topics such as people with disabilities should better be collected through dedicated surveys instead of during the population censuses, as collection of such information requires skilful enumerators with full training. Since the temporary field workers recruited for censuses/by-censuses are students, they will have difficulty in acquiring sophisticated interviewing skills from basic training.

6. Apart from the population censuses/by-censuses, C&SD has collected data on a wide range of special social topics about specific population sub-groups via the Special Topic Enquiry (attached as supplementary questions to the General Household Survey) and the Thematic Household Survey. Statistical reports on the findings are released periodically to the public. For example, a Special Topic Report was published on 22 December 2008 on the findings of a survey on people with disabilities and chronic diseases. A survey on the same topic was also conducted in 2000.

7. It is more appropriate for large-scale population censuses to aim at providing benchmark data for studying the direction and trend of population changes. In general, information on specific sectors of the Hong Kong population is collected through various dedicated surveys on the relevant special topics.

(d) Information on Hong Kong residents staying or working in the Mainland, and children born in Hong Kong to Mainland parents

8. Data relating to movement of people to and from the Mainland for work, study and retirement will be collected through the population census if the individuals concerned or their family members have a regular residence in Hong Kong and they are present to provide the information during visits in the census. People who normally reside outside Hong Kong will be categorised as non-Hong Kong residents for statistical purpose, and most of them cannot be reached during the census.

9. In respect of information on children born in Hong Kong to Mainland parents, C&SD conducted the first round of survey on Mainland babies in 2007 through interviewing their parents at the Births Registries to gather information about their plans for the newborns to reside and/or study in Hong Kong. A second round of survey was conducted and completed in February 2009 to gather the latest information on the plans of Mainland parents for their newborns.

(e) Quality assurance measures

10. The following measures have been/will be taken in the 11C to ensure quality of the collected data and published results:

Pre-census operation

- C&SD has conducted/will conduct extensive user consultations on the various aspects of design (e.g. data topics, questionnaire design, classification) to ensure that data collected in the 11C are relevant.
- All operational procedures and questionnaires will be tested thoroughly in a rehearsal scheduled in around mid-2010.
- C&SD will update the Frame of Quarters to provide a complete listing of quarters in Hong Kong thus ensuring full coverage of 11C.
- C&SD will prepare the sample design of the detailed enquiry in 11C (i.e. using the Long Form) based on scientific principle so as to obtain reliable estimates for general planning purposes.

Census operation

- C&SD will provide adequate training and guidance for the temporary field workers who will have to pass an examination conducted by C&SD before appointment.
- Field checkers will re-visit a sample of the completed assignments for checking quality of the work done and initiating immediate enhancement measures if necessary.

Initial questionnaire check

- Editors will carry out an initial check on the completed questionnaires and inform the enumerators concerned for prompt follow-up on dubious information.
- C&SD will use computer data capturing techniques to scan data of all questionnaires to minimise human input errors.
- For e-questionnaires returned, a data processing system will be in place to identify any missing or dubious entries for prompt follow-up with the respondents.

Data validation

- To ensure data quality, C&SD will use computer programs to validate the collected data comprehensively and verify dubious data with respondents if necessary.

Post-census work

- In line with international recommendations, permanent field staff will conduct a post enumeration survey to provide an independent objective assessment of the census results.
- Before release, the census results will undergo detailed statistical analysis and comprehensive professional review, including reconciliation with other related data sources and comparison with corresponding estimates from other population and social statistical systems in C&SD.
