United Nations International Seminar
on Population and Housing Censuses:
Beyond the 2010 Round
27-29 November 2012
Seoul, Republic of Korea

SESSION 8: Demands for and Provision of Census Micro Data

2011 POPULATION CENSUS OF HONG KONG
INTERACTIVE DATA DISSEMINATION SERVICE

Prepared by

Mr. Hon Kwan LAM
Census and Statistics Department
Hong Kong, China

INTRODUCTION

It has been an established practice in Hong Kong since 1961 to conduct a population census every ten years and a population by-census in the middle of the intercensal period. The aim of conducting population censuses is to obtain up-to-date benchmark information on the socio-economic characteristics of the population and on its geographical distribution. 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. Such information is vital to the Government for planning and policy formulation, as well as to the private sector for business areas.

2. In Hong Kong, the latest round of Population Census was conducted in the period from 30 June to 2 August 2011. The summary results of the 2011 Population Census were announced in February 2012. In addition, over 250 main statistical tables have been released for viewing and downloading free of charge on the website of the 2011 Population Census and more tables will be made available by phase.

3. This paper outlines the Interactive Data Dissemination Service (IDDS) of the 2011 Population Census of Hong Kong and the major issues considered in the development and use of the IDDS.

DISSSEMINATION OF CENSUS DATA IN HONG KONG

4. Based on the past population censuses/by-censuses, the following factors have been taken into consideration when determining the scope and form of census data dissemination in Hong Kong, viz.
   (i) Usefulness of data;
   (ii) Convenience and ability of users to use and interpret the statistics properly;
   (iii) Technology advancement;
   (iv) Data confidentiality;
   (v) Public awareness and accessibility of users.

5.   Consultations have also been made with major users including government bureaux and departments, academics and other organizations for preparing the data dissemination plan, which proposed products and services to be delivered, release schedule and list of core tabulations so that the census outputs could meet the needs of users as far as possible.

6.   To meet the divergent needs of users of population census data, the Census and Statistics Department (C&SD) of the Hong Kong, China adopts a multi-pronged strategy to disseminate census data for different user groups, emphasizing the service principles of friendly access and prompt delivery.

7.   Major data dissemination modes for population census data include:

(i) *Press conference* – to announce the summary findings formally so as to draw attentions of the general public and mass media.

(ii) *Analytical publications and feature articles* – to provide in-depth information on selected themes targeting major population subgroups like the youth, the elderly, ethnic minorities, or vital social issues like income disparity.

(iii) *Main tables and district-based fact sheets* – to convey the main results for the whole territory through a set of static statistical tables in spreadsheet format and summarise the demographic and socio-economic profiles of individual districts defined at various geographical levels in the form of fact sheets.

(iv) *Interactive visualisations* – to portray selected findings on various topical issues (e.g. major changes in the population age-sex structure over time) in a colourful and vivid manner by leveraging on the latest dashboard technology.

(v) *Tailor-made tabulations* – to produce tailor-made tabulations at a charge upon request by external parties.

(vi) *Access to micro-data* – to provide government users and the academics with self-help tabulation services by accessing the full dataset at the premises of C&SD or to provide them with sample datasets for off-site use for research purposes.

The C&SD has been using the Internet as the main delivery channel since the launch of an online bookstore in October 2001 and the adoption of a free download policy for all softcopies of publications and statistical tables in June 2006[1].   In particular, dedicated websites have been set up for the 2006 Population By-census and 2011 Population Census for dissemination of population census data.

8.   In terms of technology used for census data dissemination, the census products in Hong Kong have been evolved from pre-defined statistical tables and CD-ROM bundled with GIS software to facilitate spatial analysis in early 1990s to the more dynamic modes nowadays. Specifically, we have rolled out two interactive applications on our website in 2008 with the aid of GIS mapping tool to cater for different segments of users.

9.   For the general public, the "Statistics on Map" dashboards offers users better visualization through presenting statistical data in a vivid manner[2].   For the more sophisticated users, the online Interactive Data Dissemination System (IDDS), which was implemented on C&SD's website in mid-2008, allows them to build statistical tables, charts and maps of the 2006 Population By-census according to their own specific requirements through a simple and user-friendly step-by-step process on the Internet.

---

[1]   Before the free download policy, softcopies of publications were priced the same as the hardcopy counterparts for sale through the online bookstore.

[2]   A new dissemination approach is adopted in the 2011 Population Census where a number of dashboards with interactive features are developed to replace "Statistics on Map".
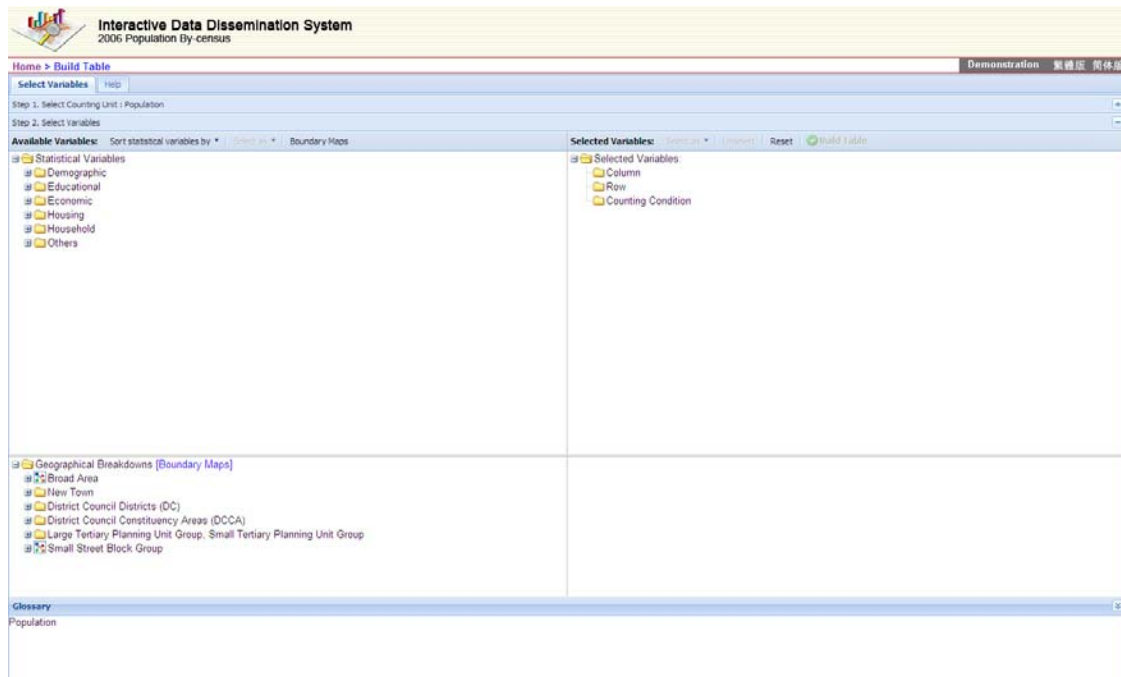
Figure 1: Interface of the old IDDS

10.   The major features of the IDDS included (i) choosing the required statistical measure and the analytical dimensions for tabulation; (ii) building a table by specifying how the dimensions are classified and displayed; (iii) customisation of the table built (e.g. adding percentages, sorting and layout modification); (iv) generation of chart and/or thematic map as a visual aid for further refinement; and (v) export of the table built in various formats for extended uses.   The IDDS tools enabled users to perform more exploratory analysis on socio-economic data online.

11.   Unlike traditional data dissemination practices, IDDS users are given autonomy in selecting variables and performing customisation based on their own data needs.   More importantly, users can directly build statistical tables and charts on the IDDS free of charge over the Internet, instead of rendering additional use of other software with downloaded data in personal computer.
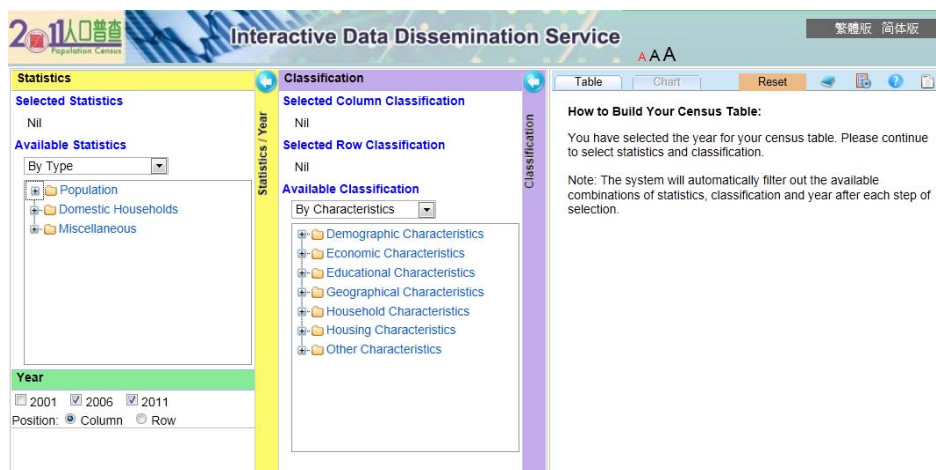
Figure 2: Interface of the enhance d IDDS

(http://itable.censtatd.gov.hk/UI/Report/Report.aspx?lang=en-US)

ENHANCED VERSION OF IDDS FOR 2011 POPULATION CENSUS IN HONG KONG

12.   With the rapid advancement in web technologies and increasing applications of population census/by-census data in both the Government and private sector, there are growing expectation of users on the IDDS towards increased data provision, dynamic data retrieval capabilities and table building functions in a more timely, interactive and user-friendly manner.   The conduct of the 2011 Population Census in 2011 gives C&SD an opportunity to further enhance the IDDS so as to disseminate the data of population census/by-census in a more user-centric manner and in response to demand for wider scope of official statistics at more detailed level.

13.   The enhanced IDDS allows greater flexibility by allowing users to produce statistical tables of the 2011 Population Census in accordance with their own data specifications (up to 2 types of statistics variables and 5 classification variables).   It also provides a better user friendly interface to build statistical tables and charts more easily.

14.   After taking into account the feedback of IDDS users, the following major aspects of the IDDS were enhanced for the 2011 Population Census:

(i)   **Data enrichment** – in view of resource constraints and the use of marco-data files for storage, there were more limitations on the tables which could be built in the early IDDS.   With a new feature of breaking down a giant cross-tabulation into various macro-data files to minimise the data input efforts, the enhanced IDDS has an extended data coverage and allows users to produce tables in accordance with their own specifications up to 5 dimensions on 2 types of statistical measures;

(ii)   **Support for temporal analysis** – with the rapid advancement in storage, retrieval and processing of data, concurrent access to population statistics from several rounds of population censuses/by-censuses becomes feasible.   Not only the data of 2011 Population Census, the enhanced IDDS includes also the data of the 2001 Population Census and 2006 Population By-census to facilitate temporal comparisons on the characteristics of population and in-depth analysis on the changes of the profile of the population over time;

(iii)   **View as you click** – the enhanced IDDS offers a new user experience in table building[3].   An updated statistical table is displayed on the right panel immediately after each selection is made on the left or middle panel in the same webpage[4].   This one-page interface design enables users to view the content and interact with the system to make further selection concurrently;

---

[3]   In the old IDDS, data users were required to select the statistical measure, the dimensions and the classification for each dimension before a statistical table is displayed for further customisation.

[4]   The basic web page of the IDDS is divided into three separate panels.   The left panel is known as the Statistics panel in the IDDS for data users to specify the statistical measure and the time dimension.   The middle panel is known as the Classification panel for data users to specify the analytical dimensions other than the time dimension.   The right panel is known as the Table panel for displaying the content of the statistical table in accordance with the selections already made.
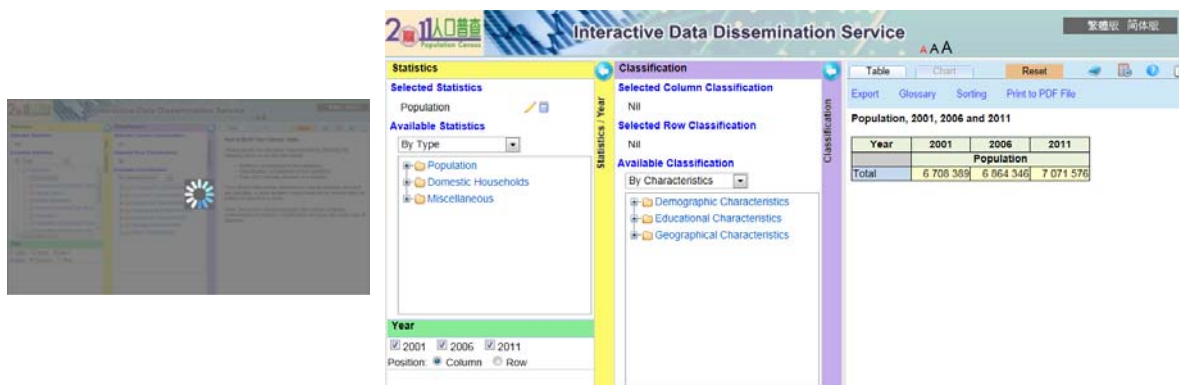
Figure 3: Concurrent display of user's selection of statistics/ classifications.

(iv) **More user-friendly** – the table building process is simplified so as to reduce the number of clicks. Users are no longer required to specify the classification for each newly added dimension. Instead, the system will automatically provide the most detailed classification available (conditionally on the last data specifications) as the default option. In addition, a number of commonly used population/household subgroups (e.g. persons studying full-time in Hong Kong) are defined as statistical measures for direct selection instead of requiring users to do so by filtering out selected classification codes after building a statistical table (e.g. a table on population by school attendance and place of study to define persons studying full-time in Hong Kong). While an animated demonstration and a user guide are provided to assist users to cope with the advanced table manipulation functions such as filtering and sorting, new users are expected to be able to use the IDDS to build statistical tables without any prior knowledge and technical know-how;
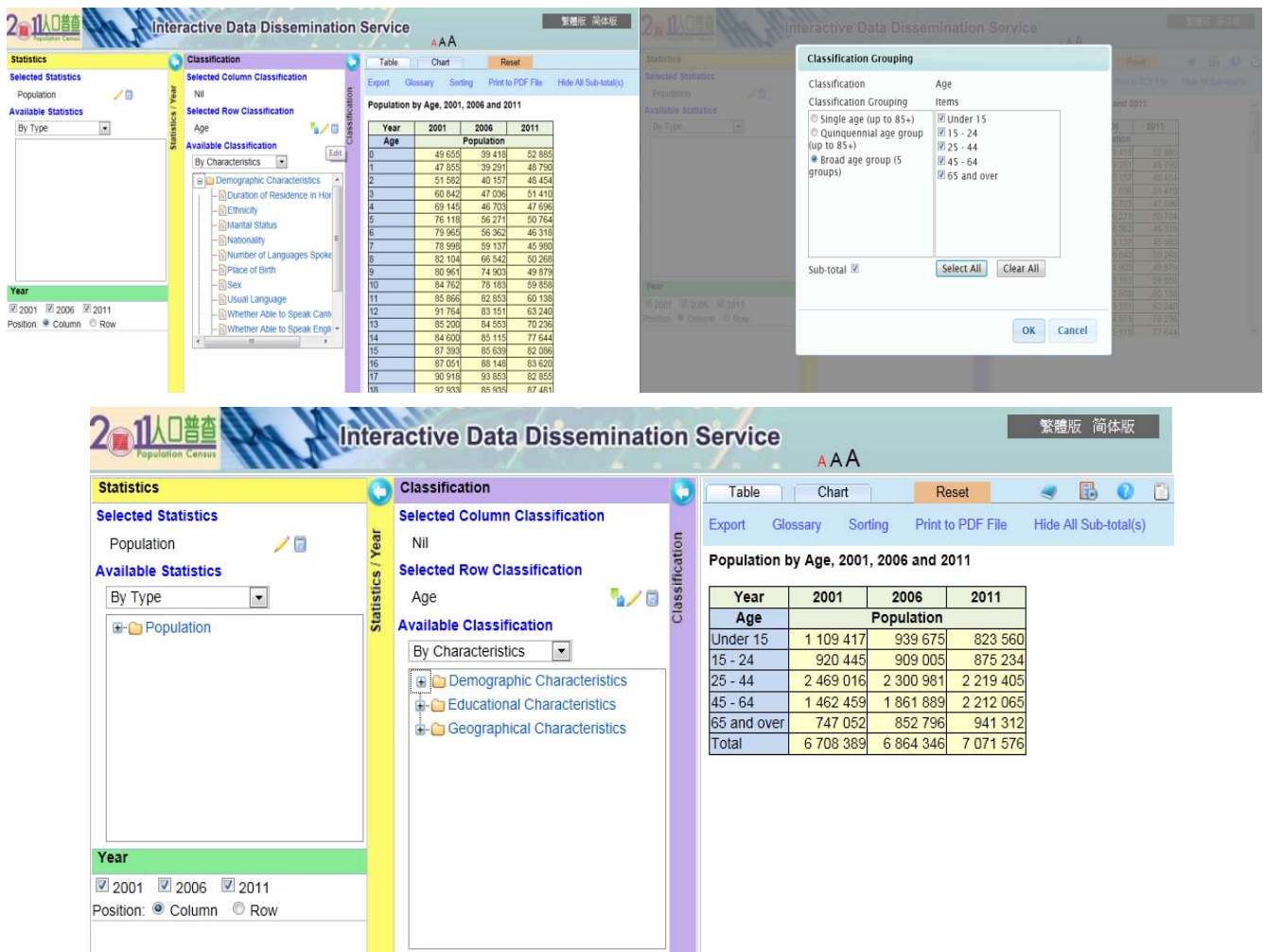
Figure 4: The most detailed classification (e.g. single age groups) is provided by the system as the default option and another classification (e.g. broad age groups) can be used if needed.

(v)　**More flexibility** – in the old IDDS, if there was one dimension with a detailed classification specified (e.g. single age groups for the age dimension), many of the other dimensions would become not available for further selection although some of them could be cross-tabulated with the former when a less detailed classification (e.g. broad age group) was used. The enhanced IDDS is designed to cater for this situation. Users are now allowed to add up to 5 dimensions so long as this combination of dimensions is feasible and the system will handle the classification hierarchy automatically by using a less detailed classification for some of the existing dimensions.

MAJOR ISSUES FOR CONSIDERATION IN DEVELOPING THE IDDS

15.   Launching the enhanced IDDS for the Population Census extended C&SD's interactive dissemination of data from population census/by-census.   By overcoming difficulties in handling a huge database and developing capacities for large demand in compiling multiple-way tables, C&SD obtained valuable experience in developing the interactive data dissemination system.

16.   The major issues considered in developing the IDDS in Hong Kong are as follows:

(i) **Participation of major stakeholders** – the success of the IDDS would hinge on the contributions of major stakeholders – users, IT experts and statisticians.   Through consultations with users (both from other government departments and from the private sector) and collection of their feedback on the use of the IDDS, the IDDS could be designed in a more user-centric manner. The major users should also be given opportunities to use the system on a trial basis before the launch of the system so that further finetuning of the system could be made.   While the IT experts have worked out solutions to develop the web-based IDDS by making use of new technologies, statisticians who are responsible for the conduct of the population census can ensure the production of detailed statistics efficiently for more in-depth and proper understanding on the social and economic characteristics of the population;

(ii) **System based on macro-data or micro-data - Balance between meeting user requirements and protecting data confidentiality** – to prevent possible disclosure of individual data in very detailed cross-tabulations, only aggregate statistics compiled on the basis of a very small number of observation units were not presented in the publications of past censuses/by-censuses.   In line with this, instead of micro-data files containing individual data records, only macro-data incorporating summarised statistics are stored in the IDDS.

The macro-data approach also helps reduce the response time as the table is built by retrieving the data from the macro-data files instead of calculating the cell values from the micro-data files. To cater for data confidentiality, data precision and system workload due to too-detailed customisation of data, the enhanced IDDS limits the selection up to 5 dimensions and 2 statistical measures. Based on past experiences, a 5-way statistical table should meet the needs of most users of population census data ;

(iii) **Ease of use versus system complexity** – there is always a trade-off between user-friendliness and functionalities. Among the five IDDS features (see paragraph 10 above), it is considered that selection of statistical measure and dimensions is of upmost importance. As a result, a lot of emphasis has been put on how to make the selection process easier and more user-friendly during the early design stage. This is also the reason why a one-page user interface consisting of three different panels is designed. However, a few advanced system functions available in the old IDDS (e.g. sorting by row[5] and suppression of empty row or column) are sacrificed as they are not suitable to the current design. In addition, users are not allowed to make any classification changes (e.g. merging cells of adjacent rows). For IDDS, simplicity is beauty on most occasions.

(iv) **Flexibility for further data expansion** – in view of the increased demand for access to wide scope of official statistics at more detail level in an interactive and flexible manner, the design of the IDDS should be scalable and flexible for further expansion. The enhanced IDDS should be able to be suitably re-deployed to support other themes of statistics;

(v) **Importance of comprehensive system testing and workload estimation** – to prevent web traffic jam, appropriate parameters with buffer should be adopted to decide the capacity of the system during system design. The system has to be tested thoroughly before its launch. Nowadays, any interruption in interactive public services may have long-term impact on the

---

[5] Sorting by column is still supported in the enhanced IDDS.

organizational images which may not be able to recover within a short period of time.   System load tests should be performed several times before launch of the IDDS to ensure smooth function of the system.

17.   The enhanced IDDS has improved the flexibility of selection variables, where both the depth and coverage of statistical tables are enriched and achieved data comparisons between different rounds of Population Censuses and By-census.   The enhanced IDDS has also marked a breakthrough for the general public's access of interactive official statistics via the Internet with a simple and user-friendly interface accessible anywhere and anytime for free, which created opportunities for C&SD to expand the coverage of IDDS to other themes of official statistics in the future.   The IDDS based macro-data instead of micro-data has also balanced user requirements and protection of data confidentiality.

18.   Since the launch of the enhanced IDDS in mid-September 2012, C&SD recorded more than 2 000 visits, generating more than 23 000 statistical tables in total.   To encourage higher usage of the IDDS, promotion will be conducted.   C&SD will introduce the IDDS through various channels, including regular publications to the public, as well as briefing to local and international users in various occasions.   C&SD will continue to monitor user feedbacks on the use of the IDDS and introduce improvements, if necessary.   Consideration will also be given to expanding the coverage of data included in the IDDS to other themes of official statistics.

END