

United Nations Statistical Institute for Asia and the Pacific

**Seventh Management Seminar
for the Heads of National Statistical Offices in Asia and the Pacific
13 - 15 October 2008, Shanghai, China**

**Managing Issues addressing the Challenges of Using Administrative Data
for Statistical Purposes**

Country : Singapore

Introduction

1 It is observed that increasingly, administrative data have become an important source of data to the statistical agencies worldwide. This is not surprising; in fact, several factors underline the gradual substitution of administrative data for data collected by surveys:

- i) the use of administrative sources helps to reduce respondent burden; and
- ii) the need to have more timely and detailed data.

2 As with most statistical agencies in the world, administrative data have increasingly become an important input to the operations of the Singapore Department of Statistics (DOS). The usage of administrative data covers both individuals and businesses.

3 This paper documents some of the main uses of administrative data by DOS for statistical purposes, followed by a discussion on the issues arising from the use of administrative data in terms of the coverage, definitions, confidentiality and quality of data etc, as well as the relationships with the administrative agencies.

Uses of Administrative Data for Statistical Purposes

4 Generally, administrative data are used for the following statistical purposes in DOS:

- i) direct tabulation;
- ii) used as proxies in estimation;
- iii) construction and updating of survey frames;
- iv) evaluation of surveys;
- v) replacement of data collection; and
- vi) editing and imputation of data.

Population Census

5 With the acceleration of the computerisation programmes in the public sector in the 1980s, there was a significant increase on the use of administrative databases in Singapore within the public sector, e.g. for school registrations and government flat applications. Data from selected public registers were first used on a large-scale basis in the Population Census 1990.

6 The improved IT capabilities for linkages between various statistical systems and administrative sources in the mid-90s led to the development of an integrated database system, known as the Singapore Department of Statistics' Household Registration Database (HRD). The HRD formed the core database on population for the conduct of the register-based Census 2000. It provided the basic count of individuals and overall profile of the population.

Business Data

7 Based on administrative data, DOS has created a business register where information on enterprises has been stored. The register is being updated at regular intervals with a constant flow of information from administrative sources. Examples of administrative data which have been used include tax information from the Inland Revenue Authority of Singapore (IRAS) as well as information on registration from the Accounting and Corporate Regulatory Authority (ACRA).

8 Basically, if a company gets itself registered at the ACRA or files in tax returns with the IRAS, then it is deemed as active and this information would be updated in the business register. In addition, for companies which are not covered in any other surveys, information on the industrial classification of such companies would also be obtained from the ACRA to update the register. Current information on business enterprises are then used to construct the statistical frame, determine the population size of enterprises, and also impute for non-response cases in establishment surveys.

9 Besides being used to create and update the business register, administrative data such as information on the GST supplies from IRAS were also use to supplement data collected from surveys for the small and medium cohorts. Since the mid 2007, these administrative data have been used to estimate the total employment and turnover statistics of start-ups at broad sectoral level.

Economic Data

10 While the foundations for measuring the services sector in Singapore have been firmly laid, there is still the additional need for timely, accurate and reliable short-term economic indicators to assess the economic performance of this sector.

11 In developing these indicators, DOS has relied as far as possible on administrative data which has the advantages of being available on a high frequency, more comprehensive and less labour-intensive than surveys. A wide range of short-term indicators is derived from administrative by-products arising from the operations of several government agencies. Some examples:

- i) monthly information on external trade are used to assess the performance of entrepot and wholesale and retail trade;
- ii) monthly data on room days occupied, visitor arrivals, average occupancy rate, are used to assess the performance of hotels and restaurants.

12 On top of that, data collected from the Population Censuses had also being used to cross validate data captured under other major surveys as well as to serve as a gauge for estimating the output of activities not covered in any of the major surveys.

Management Issues Arising from the Use of Administrative Data

Coverage

13 Data in administrative sources have generally been collected for a specific administrative purpose, and the needs and priorities relating to that purpose are unlikely to be the same as those of the statistical system. The variables measured and tracked by the administrative agency tend to be rather different from those variables and categories which are most suitable for statistical purposes.

14 For instance, some fields required for statistical compilation (e.g. 'turnover') in tax forms are not required by the IRAS for tax assessments and hence they are not made mandatory. As a result, some taxpayers may choose not to provide information pertaining to these fields which are important inputs for statistical estimates.

15 Where there is not enough administrative data, DOS will try to supplement these with survey data. Of the 54 items collected for the Population census 2000, only 8 items could be obtained directly from the HRD; the remaining items were collected from a 20 percent sample of the population. In the development of short-term indicators required to assess the economic performance of services sector, additional information were collected from higher frequency surveys such as the survey of *Quarterly National Income Estimates*.

Definitions and Classifications

16 The definition and classification systems used within the administrative sources may also be different to those used in the statistical system. It is important to maintain consistent definitions and classifications between the administrative sources and the statistical institutes if administrative data were to be used as inputs to statistical operations. In this aspect, DOS has taken the lead as the national statistical coordinator in developing national standards on the occupation, industry and education codes through inter-Ministry consultations and also in promoting the adoption of these standards by the various government agencies.

Matching of Data

17 If data from several administrative sources are used, it is likely that the statistical agencies will be faced with problems matching the data. Matching will be relatively easy if there is some form of common identification number. The HRD had been built by matching various government databases containing administrative information on individuals and this was made possible because every resident in Singapore had been issued with a personal identification number (UIN) and every foreigner staying or working in Singapore needs to hold a pass with a unique foreigner identification number (UFN).

18 From 1 January 2009, all entities registered in Singapore, such as businesses, local companies, limited liability partnerships (LLPs), societies, representative offices, healthcare institutions and trade unions, will have a unique entity number (UEN). This is an initiative under the Singapore government ICT master plan. As the UEN will be used for all correspondences and interactions with government agencies, it will enable data sharing across government agencies, as well as facilitate data matching across the various administrative databases.

Data Confidentiality

19 Even with the legal authority to access administrative data, careful consideration has to be given to the public perception of the use of administrative data beyond their original

purpose. There may be public unease at the thought of data being shared around government. Insofar as data confidentiality is concerned, all administrative data acquired by DOS are protected under the Statistics Act. ie only confidentiality-protected aggregate data are provided back to the agency or release to the general public. This is the same practice as the statistical institutes of most countries.

Quality of Statistics Derived from Administrative Data

20 It is necessary to look into the overall quality of the administrative data. It is not always the case where administrative data are superior to the survey data collected. In such cases, survey-based data are used to benchmark administrative-based estimates.

21 Generally, DOS evaluates the quality of administrative data through the following ways:

- i) cross-checking against available survey data and other administrative sources;
- ii) performing outlier detection within the appropriate strata; and
- iii) comparison with data of previous periods at enterprise level.

22 When conducting the Population Census 2000, checks were performed to ensure that the quality of records stored in the HRD, which is the core database for the Census, were of high standard. Systematic error corrections were carried out for the database; comprehensive validation and verification rules were built into the database's ongoing updating cycles and individual records which failed the validation checks will be re-verified and corrected.

23 Comparisons with official population estimates were also done, for example, population figures based on HRD records were generated and compared with those obtained from General Household Survey (GHS). Where there are differences, possible explanations are studied and corrective actions were taken subsequently.

Relationship with the Administrative Agencies

24 With the growing use of administrative records, statistical agencies are becoming increasingly dependent upon other agencies for the uninterrupted flow of input data to their statistical programs. It is therefore necessary and important for the statistical institutes to maintain good relationships with the administrative source agencies and also to establish close co-operative arrangements with them.

Training and Data Sharing

25 In order to understand better the administrative data collected by various agencies, DOS will communicate directly with these agencies or convene meetings to clarify on the scope, characteristics and potential quality issues of the data provided. Some of these agencies may also conduct courses sharing general information on the data collected but most of these courses were not able to fully cater to DOS's requirements as our questions are more specific pertaining to the use of data.

26 Periodically, DOS conducts courses on Singapore's industrial and occupational classifications as well as sharing sessions on subject matters so that the source agencies

would be more aware of the kind of work done by DOS and hence appreciates the need for more timely data for statistical purposes.

Formal Agreements with Administrative Agencies

27 Administrative sources are generally set up for the purposes of operationalising and monitoring government policies. If a policy changes, administrative sources may be affected in terms of coverage, definitions etc, or possibly even abolished completely. Such changes would have an adverse impact on the statistics compiled based on administrative data. .

28 These risks have been managed by additional provisions in the government instruction manual (IM). In June 2007, the IM's information management policy on data administration was amended which states that agencies whose administrative data are currently used by DOS for statistical compilation should inform DOS of any plan or intention to review their data collection or data processing systems, so that statistical considerations could be factored into the changes.

29 In addition, agencies should also inform DOS in advance of the introduction of any new administrative data, systems or processes arising from new government policy or scheme so that considerations for the collection and use of new administrative data for statistical compilation could be factored into the new systems or processes.

Major Obstacles/ Weaknesses Faced

30 Even with the change on the policy on Data Administration, some agencies have dropped certain data items from their administrative systems in the interest of lowering business/compliance costs. Such data items are critical for statistical compilation; In general, DOS accepts the administrative data as it is. Some agencies also require detailed justification for each data item requested and DOS also needs to pay cost recovery fees.

Conclusion

31 Administrative data are increasingly becoming an important source of statistical data for DOS. While we acknowledged that there are many benefits in using administrative data for statistical purposes, we are also mindful of the many issues arising from using such data, especially when the objectives of collecting administrative data are often not for statistical purposes.

32 We understand that it is important to have detailed knowledge of the administrative data, for example, the main purpose of collecting such data, the way the data have been collected and also the way they have been processed. Only through thorough understanding of the source will we be able to have a more accurate assessment of the strengths and weaknesses of the administrative data and how they could be applied for statistical purposes.

33 As such, DOS has been conscientious in establishing a good rapport with other government agencies. We have been working very closely with them to develop a central directory on government data so as to leverage on the collective statistical knowledge of all government agencies. DOS has also been diligent in assessing the quality of administrative data acquired and along the way, had resolved some of the issues which arose as a result of using such data.

34 As the demands for statistics on all aspects of lives continue to grow in a climate of tight budgetary constraints, administrative data would be an alternative source for statistical purposes for DOS.

Singapore Department of Statistics

October 2008