

11th Meeting of the Heads of National Statistical Offices, 6-8 November 2006, Tokyo

Topic 2: Improvement of data collection for basic economic and social statistics

**Title: Establishment Directories and Economic Databases in Malaysia:
Potential benefits and challenges in data sharing**

by

Aziz Mohammad*

Abstract

Collaborative use of data from other agencies could substantially reduce NSO's operational cost of data collection. Since secondary data and administrative records are used in updating both the establishment frame and the economic databases, efficiency in accessing and updating is of primary importance. Thus, tremendous gains from data sharing arrangements could be achieved through on-line linkages between agencies and the establishment of common national databases. In the case of Malaysia, the Department of Statistics Malaysia has succeeded in establishing an on-line networking system with its branch offices for the purpose of updating the establishment frame, while provision for linkages with other agencies has been included in the department's ICT Strategic Plan which is expected to be implemented over the course of the Ninth Malaysia Plan, 2006-2010. However, there are challenges along the way and the paper highlights the department's experience in collaborating with other agencies concerning data on SMEs and the agriculture sector. The rules and regulations governing individual agencies limit the sharing of information. Data from various agencies are stored in different formats and are subjected to different classification and coding systems. Frequency of data collection also differs between agencies and this has also affected sharing and wider usage of such data.

1. Introduction

Advancement in ICT is expected to result in tremendous gains from data sharing arrangements through on-line linkages between agencies and the establishment of common national databases. With proper co-ordination, primary data collection activities by a National Statistical Office (NSO) need to be confined only to data items not covered by other agencies.

However, from the experience of the Department of Statistics, Malaysia (DOSM), such benefits are not expected to be forthcoming in the immediate future. Tracing DOSM's previous efforts from developing the Central Registry System (CRS), to its more recent endeavors in establishing the small and medium scale enterprise (SME) and agriculture databases, it is apparent that besides technology, the success of such collaborative efforts depends to a large extent on the commitment from all the relevant agencies.

* The writer is the Deputy Chief Statistician, Macro & Development Programme, Department of Statistics, Malaysia.

2. Establishment Directories of DOSM : the CRS

The directory of establishments and enterprises of the Department is maintained in the Central Registry System (CRS). The CRS was set up in 1994 in a PC based system using dbase IV and was upgraded in 1996 into a Local Area Network (LAN) Environment. In 2002, the system has been upgraded into the Wide Area Network (WAN) Environment. It was also redesigned to take into consideration the additional needs of users, the requirement for new survey frames and to accommodate the new MSIC codes.

The main CRS profiles comprised of Enterprise, Establishment, Operational Work Control (OWC) and Directories. Enterprise and establishment profiles serve as a data base of all enterprises/establishments to be covered by the economic censuses/ surveys of the Department. The OWC profile provides a facility to control data collection operations conducted by DOSM's field officers. These profiles are being linked to avoid occurrence of duplicate records as well as to co-ordinate maintenance and data entry. The directories contain lists of companies registered with the Companies Commission of Malaysia (CCM) together with post codes, enumeration block codes, and industry codes. These directories are being linked to both the enterprises and establishment profiles, and are used mainly for quality control checks. Prior to the inclusion of the SME census frame, CRS contained 121,800 establishments, most of which are big establishments.

CRS is a system that manages business establishments frame for DOSM. The purpose of CRS is to provide a profile of business establishments and enterprises that are used as sampling frames of the related surveys conducted by the Department. The profiling data consist of data elements that are necessary for locating the sampling unit in the survey such as Reference Number (ID) Name of Establishment/Enterprise, Address, Operational status, (active, dormant, closed etc), Activities, and Data required for Sample Selection (employment, revenue, etc). The database also contains auxiliary data elements needed for stratification in the sampling design such as the type of activity of each business establishment, the number of employees, the value of production/sales, the value of expenditure and assets.

The system consists of a number of modules that will generate a listing of establishments based on query, perform database maintenance functions, and provide reporting facilities and print labels. The system is deployed to all state offices via a Wide Area Network (WAN).

After merging with the SME Census Frame in 2004, the CRS database contains 1.7 million records of establishments from various economic sectors. The records were originally compiled from secondary data obtained from two government agencies, namely the Registrar of Businesses (ROB) and Registrar of Companies (ROC) - both are now under the Companies Commission of Malaysia (CCM). Subsequently, in the course of their involvement in building

the SME census frame, the following agencies have shown interest to provide the relevant information for enhancement of the CRS database:

- i. Employees Provident Fund (EPF);
- ii. Inland Revenue Board (IRB);
- iii. Construction Industry Development Board (CIDB);
- iv. Agriculture Bank of Malaysia;
- v. Credit Guarantee Corporation (CGC); and
- vi. Veterinary Department and Fishery Department.

New records as well as updated particulars of existing records from the CCM are regularly merged with the existing CRS database.

3. National SME database

3.1 National SME Development Council (NSDC)

Realizing the importance and potential contribution of SMEs, the NSDC was established in June 2004 as the highest policy making body concerning the comprehensive and coordinated development of SMEs across all sector of the economy in Malaysia. The role of the council is to provide the strategic direction for government. policies on SME development so as to ensure a more competitive and resilient SMEs.

Among the initiatives announced are the formulation of broad policies and focused strategies for the development of SMEs across all sectors; the expansion of development support programmes and facilitating access to financing; the adoption of specific and standard definition for SMEs according to economic activities; and the establishment and maintenance of comprehensive National SMEs database.

3.2 SME Database

DOSM has been entrusted with the responsibility of creating a comprehensive national SME database based on the agreed definition of SME. Relevant and timely information on SMEs, such as their operational and financial status as well as their needs for further advancement, is essential both in the formulation of the appropriate policy in promoting SME development and for the subsequent monitoring of developments in the SME sector by the relevant authorities.

For the purpose of creating the National SME Database, an action plan has been drawn up for subsequent implementation.

The action plan includes;

- i. Conducting a nation-wide Baseline Census of SME establishments and enterprises covering the main economic sectors of manufacturing, services and agriculture;

- ii. Establishment of a frame/database which includes data harmonization and integration between the various providers of SME information.

Implementation of the action plan would thus involve the formation of the SME database steering committee; resolving harmonization and integration issues; setting timelines for conducting the baseline census; and monitoring of development in the implementation of the SME database.

3.3 SME Census Frame

The SME Baseline Census was conducted in 2005 with 2003 as the reference year.

The tasks undertaken in preparing the Census frame include:

- a. obtaining a list of all companies and businesses registered with Companies Commission of Malaysia (CCM) in the selected sectors;
- b. establishments/enterprises not registered with CCM will be sourced from relevant Ministries and agencies involved in SME development; and
- c. Data cleansing and screening.

For the construction of the Census frame, CRS was used as the initial source of information on SMEs.

The additional sources for the frame included all establishments and entrepreneurs registered with the Ministries and related agencies such as Ministry of Agriculture and Agro-based Industry, Ministry of Human Resources, Ministry of Entrepreneurs and Co-operative Development, Central Bank of Malaysia, Employee Provident Fund (EPF), Inland Revenue Board (IRB), Small and Medium Industries Development Corporation, Credit Guarantee Corporation, and the Local Authorities.

The list of establishments and entrepreneurs obtained from the agencies mentioned above totalled 4.6 million. The various lists had to be screened, matched, verified and updated to eliminate duplication.

From a total of 4.6 million, the Census frame finally comprised a total of 1.73 million establishments/entrepreneurs in the manufacturing, services and agriculture sectors.

3.4 Database Structure

Data from the Baseline Census served as a starting point for the subsequent input to the SME database. SME database is expected to cover a wide range of small and medium scale enterprises, including non-registered establishments. Hence, the SME database contains records of establishments from various agencies such as CCM, EPF, and IRB.

SME database consists of more than 200 data fields pertaining to 550,000 active establishments enumerated in the SME Baseline Census. A subset of these data fields will in turn be the main input to CRS.

The structure of the database will consist of both establishment and aggregate level data. Thus, the updating of the frame will be done through inter-censal data collection activities. For out-of-sample establishments, screening process should be done to get the latest information of the particular establishments.

4. Agriculture Database

The Agriculture Database is still in the early stage of development. Currently, this system has been developed based on the 2005 Agriculture Census. The database is expected to be completed by year 2007.

The Agriculture Census involved more than 300,000 agricultural households. Agriculture database was developed based on the census data which contained more than 200 data fields. A subset of these data fields will be the main input to the agriculture frame for subsequent data collection exercises.

The structure of agriculture database and the corresponding mechanism for system interaction with various agencies are expected to share some common features with that of the SME database.

5. Issues and Challenges

- i. Ideally, assigning a unique identifier for each establishment should facilitate the process of checking and matching. However, source agencies differ with respect to the way they assign identifiers to the establishments. Some sources do not provide these information while for some the particulars were incorrect/incomplete. Certain information, such as description of activity, operational status, name, address, and post codes were either not reported or incomplete. Details on availability of information from various sources are given in **Appendix 1**.
- ii. Records from source agencies were stored in different formats and are subjected to different classifications and coding systems, thus making the job of matching extremely difficult and cumbersome. Often manual checking had to be done. Frequency of data collection also differs between agencies and this has also affected sharing and wider usage of such data. Differences in the codes and formats used between DOSM and various agencies are listed in **Appendix 2** and **Appendix 3** respectively.
- iii. In looking forward to data sharing, DOSM is also subjected to time constraints. Since each agency uses a different system/software to run analysis/prepare report, this will slow down the task of developing the data sharing system.

- iv. Although using various sources to update the frame is helpful, overlapping enterprises/establishments does occur. Different sources reported different names and registration numbers resulting in duplicate elements in the frame.
- v. The link between a parent company and its subsidiary companies is not indicated in the ROC. This limitation is a disadvantage, especially in building up enterprises to the holding level in the CRS.
- vi. Legally, certain agencies are not allowed to share information. For example; LHDN is empowered by Section 138 of Income Tax Act 1967 whereby “ all the information of the tax payers are confidential and will not be divulged to third parties with the exception of written permission from the respective tax payers”.
- vii. The matching process between records from the CRS and records from other sources is done manually (using MS Access) to detect new establishments and update the existing frame by using the registration number from ROB/ ROC (if available). For those not reported according to ROC/ROB registration numbers, matching will be done using establishment names. This is a massive exercise since it is done manually and involves a large number of record.
- viii. Maintaining a sampling frame is quite a challenging job because at any time the enterprise/establishment will change its status, activity and etc. The issues are where to get the latest information besides the internet and how are we going to know which company has been taken over, closed down, merges or simply not operating. Such cases will only be known if it was published in the electronics media or newspapers, or otherwise will be detected only during the survey proper.

7. *Benefits*

The benefit from this massive exercise is the updating of CRS. Here posting of current information from the Baseline Census as well as the availability of the establishment frame for the agriculture sector augurs well for sampling purposes. This too will be extended to the other economic sectors such as construction and mining & quarrying after the completion of the currently ongoing Economic Census 2006.

Similarly, the linkage between the Agriculture Census and other data sources, most of which are also data users (**Appendix 4**) should be used to the fullest in constructing and updating the sampling frame for agriculture.

8. **Proposals for collaborative use of secondary data sources**

- i. Regarding the administrative records, there should be a harmonization in the use of common definitions, codes and classifications between data providers. Therefore, the Department should form alliances with data providers to develop shared systems that can interface with CRS.

- ii. Enhance the existing CRS to be used as a national sampling frame for establishments. The enhanced CRS should consist of a core database with the required data structure and managed by 4 system modules:
 - i) Gateway Interface module – maintains the system information of the parties that will participate in the automatic updating process, maintain access security, schedule auto update invocation and updating routines.
 - ii) Update Interface module – maintains the list of fields for each survey that supplies information for the updating process.
 - iii) Maintenance module – capable of display, printing, general system maintenance, file and text editing, data entry, query, report generation, network and connectivity setting.
 - iv) Sampling Generation module – able to display sampling stratification control variables, simulate sample size determination, produce random sample units, print address labels and maintain record of sampling units selected.

Regarding the system management of the CRS, the system needs to be revised/updated to resolve data replication issues for merging updated files from regional offices.

- iii. Since the matching process is a crucial task in detecting new births and updating the frame, a proper well planned system needs to be developed to cater for this massive exercise. Although there was a system called NESS (New Establishments Screening System), this system needs to be upgraded.

Information Available At Various Agencies

INFORMATION	AGENCY			
	CCM	CIDB	EPF	IRB
Registration Number	✓	✓*	✓*	
Establishment Name	✓	✓	✓	✓
Address	✓	✓	✓	✓
Registration Date	✓	✓		
Activity	✓			
Status	✓		✓	
Owner's Name	✓			
Closing Date	✓			
List of projects		✓		
Contractor's grade		✓		
Project Expiry Date		✓		
Telephone Number		✓		
Employment Number			✓	

Classification And Codes Used by DOSM and Other Agencies

Classification of Activity

DESCRIPTION	Agency				
	DOS	CCM	CIDB	EPF	IRB
1 Manufacture of soft drinks, production of mineral waters	15541 15542	313000 313400	-	31	15541 15542
2 Manufacturing of Clothing's	18101	322210	-	33	18101
3 Manufacture of Footwear	19200	324000	-	34	19200
4 Building Construction	45201	511000 512000	G1-G7 (B)	53	45201
5 Civil Engineering Construction	45209	510000 513000 528000 528100 528200	G1-G7 (CE)	52	45209

Classification of Citizenship

DOSM							OTHER AGENCY			
POPULATION CENSUS 2000		AGRICULTURE CENSUS 2005		HIS 2004		LFS/MIGRATION 2006		JHEO	NPFDB (LPPKN)	
01-15	State in Malaysia	01	Citizen	1	Malaysia	1	Malaysia		0	Malaysia
16	Singapore	02	Singapore	2	Singapore	2	Singapore	Singapore	1	Singapore
17	Indonesia	03	Indonesia	3	Indonesia	3	Indonesia	Indonesia	2	Indonesia
18	Philippines	05	Philippines	4	Philippines	4	Philippines		3	Philippines
19	Thailand	04	Thailand	5	Thailand	5	Thai	Thailand	4	Thailand
20	Brunei									
26	Bangladesh	06	Bangladesh			7	Bangladesh	Bangladesh	6	Bangladesh
30	Others ASEAN Country					6	Others ASEAN Country		5	Others ASEAN Country
38	Others Country	10	Others Country	6	Others Country	8	Others Country	Others Country	7	Others Country

List Of Other Government Agencies Selected To Contribute For Frame Of Establishment

Name of OGA	Frequency of Update	IDENTIFICATION			ADDRESS OF ESTABLISHMENT				CONTACT INFORMATION			BUSINESS ACT		ECONOMIC FACTORS						Database System Used		
		Reg. No	Licence no.	Estab. Name	Street's Name	Post Code	District or Mukim	State	Contact Person	Phone No	Fax No	e-mail	Activity Name	Business Code	No of Employe	Gross Income	Total Output	Total Wages	Total Export		Capital Expend.	Number of Records
National Tobacco Board	1,2	C	C	C	C	N	C	C	C	N	N	C	C	C	N	N	N	N	N	N	1,000-5,000	Relation DB
Federal Land Development Authority (FELDA)	3	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	More than 100,000	DB2 (AS400)
Department Of Fisheries	2	C	C	C	C	C	C	C	C	C	C	C	C	C	N	N	N	N	N	N	20,000-50,000	MySQL
EPF	1,3	N		C	C	C	C	C	C	C	C	C	C	C	N	N	N	N	N	N	More than 100,000	Relation DB
Pepper Marketing Board	1,2	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	5,000-20,000	MS Access
Companies Commission of Malaysia	1,3	N		C	T	N	T	T	T												700,000 - 3,000,000	
Inland Revenue Board				C	T	C	T	T													600,000	
Agricultural Bank Of Malaysia	1,2	C	C	C	C	N	C	C													More than 100,000	DB2 (IBM)
Malaysian Construction Industry (CIDB)	1,2	C	C	C	T	N	T	T													40,000	MS Access
Minerals and Geoscience Department of Malaysia	1,2			C	T	N	T	T													3,000	

Footnote-

- Frequency of Update
 1 Once during registration
 2 Once a year
 3 Twice a year
 Data Type
 C Character
 N Numeric
 T Text
 I Integer

List Of Other Government Agencies Selected To Contribute For Agriculture Frame

Ministry Of Agriculture & Agro-Based Industry
Ministry Of Rural Development
Ministry Of Plantation Industries & Commodities
Department Of Agriculture
Department Of Veterinary Services
Department Of Fisheries
Fisheries Development Authority Of Malaysia
Farmers Organization Authority
Federal Agricultural Marketing Authority (Fama)
Sabah Rubber Industry Board
Malaysian Agricultural Research And Development Institute (MARDI)
Cocoa Board Of Malaysia
Rubber Board Of Malaysia
Malaysian Palm Oil Board
Federal Land Consolidation And Rehabilitation Authority (Felcra)
National Tobacco Board