STATISTICS DIRECTORATE

**SBS Expert Meeting "Towards better Structural Business and SME Statistics"
OECD, Statistics Directorate
3-4 November 2005
La Muette Room 4**

Agenda Item 6 : Business registers and SMEs

item 6(a):

Statistics New Zealand Business Frame (BF) strategy and developments related to statistics on SMEs and the support of longitudinal business statistics

Mr. G. Mead, New Zealand

1. Executive Summary

This paper provides an overview of the Statistics New Zealand Business Frame (BF) and its use in producing longitudinal business statistics and statistics on SMEs (small and medium sized enterprises).

The longitudinal scope of the BF is currently being enhanced with the development of the Longitudinal Business Frame (LBF) using linked administrative employee and employer information. Two different techniques, one based on probabilistic matching of firm births to existing firms, and the other based on tracking clusters of employees who appear to shift between employers, are used to create longitudinal links in the LBF.

The BF is also used as a source of statistics on SMEs. This paper describes a range of issues related to the production of SME statistics from the BF. Such issues include the definition of SMEs, data quality, data availability and international comparisons.

2. Business Frame

Statistics New Zealand's Business Frame (BF) is a list of private and public sector businesses and organisations ("enterprises") engaged in the production of goods and services in New Zealand. The BF is comprehensive and is continuously maintained to reflect real world changes e.g. business start ups, businesses closures etc. The BF is estimated to cover 99 percent of business sector economic activity covered by the tax system. To be included on the BF a business has to meet criteria based on minimum levels of activity, a boundary described as "economic significance". There are approximately 420,000 live enterprises recorded on the BF.

The BF has a statistical unit model for all businesses consisting of three units:
- Enterprise (ENT)- legal unit
- Kind-of-activity unit (KAU)- an enterprise is only structured with more than one KAU if it has very significant activities in more than industry and the enterprise can supply financial information by industry
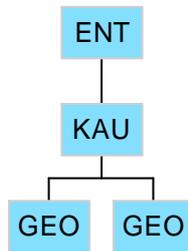- Geographic unit (GEO)- local unit

Examples of Statistical Unit Structures on the BF

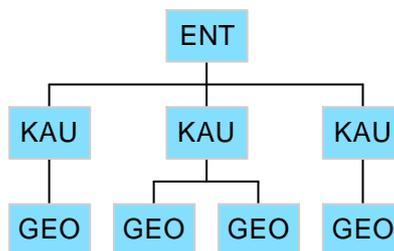A small single location business will be represented on the BF as:
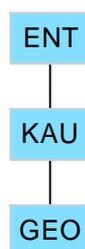
A business with two locations (geographic units) will be represented as:

```
        ENT
         |
        KAU
       /    \
     GEO    GEO
```

Whereas a large business, with significant activities in different industries that meet the criteria for more than one KAU will be represented as:

```
              ENT
          /    |    \
        KAU   KAU   KAU
         |    /  \    |
        GEO GEO GEO  GEO
```

The majority (97 percent) of businesses on the BF are single geographic unit enterprises. However 10 percent of all geographic units belong to enterprises with more than one geographic unit and 47 percent of total business employment recorded by the BF is in geographic units that are part of multi geographic unit enterprises.

The information recorded about each business on the BF includes:
- Legal name
- life cycle
- contact details
- ownership links with other enterprises, enterprise group relationships
- number of persons employed by the business
- link to tax system ID
- annual financial activity size measure (derived from VAT type tax information)
- standard classification codes for:
    industry
    institutional sector
    location
    type of business
    nature of overseas transactions

The prime source of maintenance information for the BF is administrative tax data. Processes using monthly tax data create new businesses on the BF, cease businesses and update employment size measures. For medium and large businesses Statistics New Zealand surveys collect information to confirm and supplement the tax based information. Other sources such as media reports and administrative information related to company registration, etc are also used.

Changes on the BF are recorded attribute by attribute in "history" tables. When changes are made to an attribute (e.g. industry code) or a statistical unit relationship (e.g. link from a geographic unit to an enterprise), both the date of the change on the BF and the date of the change in the real world are recorded. In this way the history tables can be used to track changes in BF variables and statistical unit relationships over time.

The BF is used by Statistics New Zealand as a:
- frame for all the business surveys operated by Statistics New Zealand
- frame work for the integration of administrative data with Statistics New Zealand data
- source of standard classifications
- source of statistics on business numbers and business demographics

3. Using the BF as a frame for longitudinal business statistics and business demographic statistics

In order to produce meaningful statistics about the economic performance and life cycle of businesses it is necessary to be able to link administratively driven changes. A firm may undergo several changes in its lifetime, in addition to a firm's birth and death. For example, legal or administrative entities close down or are created due to break-ups, mergers, split-offs, takeovers, or re-structuring. It is important that the true creation or destruction of firms, as opposed to administrative reshuffles, can be identified in a longitudinal business frame.

The Eurostat Business Register Recommendations Manual – March 2003 Revision defines birth and death as follows: "a birth amounts to the creation of a combination of production factors and a death amounts to the dissolution of a combination of production factors, both with the restriction that no other enterprise is involved" and goes on to say that "an enterprise is considered to be continued if its production factors are continued. It is discontinued if its production factors are discontinued". It also states that "A distinction must be made between the (conceptual idea of) real observable world and its reflection in administrative files and in statistical business registers".

Statistics New Zealand is progressively moving towards producing business demographic statistics based on this definition by using a range of information sources.

3.1 Current situation in 2005

The current (2005) BF maintenance processes automatically create, from administrative data, a new enterprise (ENT) unit record for each new legal unit within scope of the BF maintained population. Similarly a new KAU unit is created for every new ENT created on the BF. A new GEO is also created automatically for each new KAU. However many new legal unit births are actually "existing" businesses. For instance:

1. An existing business A assumes a new legal status as business B with a new tax registration, for example, changing from a partnership to a limited liability company. The business continues to use the existing factors of production.
2. Existing business A ceases. Some or all of its functions continue as new business B.
3. An existing business A splits off some of its functions to a new business B. A continues trading.
4. Existing businesses A and B amalgamate to form a new business C.
5. Existing business A acquires all of the production factors of business B.

While all of the situations described above involve the birth of new administrative numbers, none of them is regarded as the birth of a new business; these are all cases of administrative churn. Businesses cannot be linked over time by enterprise number or IRD number, because either new numbers have begun to be used or existing numbers have

ceased to be used. Administrative churn peaks at the end of one tax year and the beginning of the next. In New Zealand, where the most common balance date is at the end of March, this corresponds to a peak in April (Kelly, 2003).

For small enterprises no further action is currently taken on the BF. Administrative data provides all the essential BF classification and size information. Monthly processes using administrative data also automatically cease enterprises to ensure that the BF is not "duplicating" entities.

However for new medium and large enterprises a Statistics New Zealand frame update survey is used to supplement and confirm the administrative data. The survey includes questions to identify enterprises that have taken over existing enterprises or changed their legal structure. For industries where the geographic unit represents a fixed place of business (e.g. a farm, a factory, a shop) and the prior owner of the geographic unit can be identified then the geographic unit is transferred from the old owner (enterprise) to the new owner (enterprise). The unique ID of the geographic unit remains the same, the link to the owning ENT/ KAU changes. This process is not comprehensive however, as it relies on good source information and the ability to identify the historical owning business on the BF. This process is also resource intensive.

Out of 50,000 new enterprises on the BF over a 12 month period, 10,000 (20 percent) ended up being associated to an existing geographic unit(s) on the BF. The new enterprises with transferred geographic unit(s) however accounted for almost 60 percent of the total employment of all new enterprises. So the existing practice of geographic unit transfers contributes significantly to the BF providing longitudinal information for the larger employing businesses. A range of business demographic statistics on business entry, exit and growth rates has been produced using this current level of longitudinal information on the BF.

3.2 Developments under way in 2005 - Longitudinal Business Frame (LBF)

Developing a longitudinally linked business database, the Longitudinal Business Frame (LBF), is part of a strategic project at Statistics New Zealand. The Linked Employer Employee Data (LEED) project integrates existing employer and employee information to provide new insights into the operation of the New Zealand labour market and its relationship to business performance.

When BF geographic unit reference numbers change in the absence of corresponding real world events, the ability to track continuing businesses over time is lost. LEED and the LBF include processes that assist in recognising when businesses in different periods are actually the same business, even when their identifiers change in the source data. These processes are said to 'repair' the respective IRD and LBF identifiers. Both data sources contain information on changes in economic entities which can be used to infer the existence of real world events and "undo" spurious business births and deaths.

3.2.1 Creating longitudinal links using probabilistic matching

BF births are matched against existing units on the BF to identify if the birth is actually an existing geographic unit on the BF changing legal structure or owner ship. Births to the BF in a particular month are matched to the businesses that were on the LBF in the previous month. Geographic units are matched on variables common to both files, including name (both trading name and legal name), location (street address and region codes), industry code and phone number. This mix of variables contains some that are highly discriminatory (such as phone number, legal name) and others that are less so (such as street address and other location information). The matching routine processes the data in six passes. The first two passes limit potential matches to units in the same area (meshblock) and look for an exact match on legal or trading name. The third pass takes advantage of the discriminating

power of the telephone number and uses this as a blocking field together with a combination of location, industry and name matching information to identify continuing businesses. Later passes cast the net more widely by gradually extending the size of the region that units to be compared may belong to and using various combinations of location, industry, and name matching fields, as well as increasingly higher cut-off thresholds, to identify links where one or more of the fields is in error.

A linked record pair occurs when a pass brings together two records, each with a different BF number, but belonging to the same business. When a link is made the LBF is updated by recording the details of the new identifiers.

Clerical review of the probabilistic matching process showed that the level of false positives was approximately 2 to 3 percent in any month, while the level of false negatives was no more than 5 percent. Using this method, approximately 15 percent of firm births in any month are shown to be continuing businesses. At the beginning of a tax year (April), the percentage of spurious births can be as high as 20 percent. This confirms how important the use of probabilistic matching is to enhance the value of the LBF data. Statistics New Zealand's aim is to identify as many continuing businesses as possible, without ending up with too many erroneous links. Missing matches inflate the number of firm births and deaths.

3.2.2 Repairing longitudinal links by tracking employees

In the New Zealand tax system income derived from salary and wages has pay-as-you-earn (PAYE) tax deducted by the employer. All employers must file an Employer Monthly Schedule (EMS) every month to Inland Revenue (IRD). Large employers file an EMS twice a month. Thus every month employers list each employee and that employee's details. LEED uses EMS data as a base and adds employer industry and region information from the BF. In addition, information from the EMS data is used to repair longitudinal links between employers, in both the LEED system and in the LBF system.

Employees and employers are uniquely identified within the tax system by their IRD numbers. An EMS record represents an employee employed by a particular employer in a given month, including tax and earnings details. This pairing of employers and employees means it is possible to identify administrative changes by following groups of employees that change employer IRD number in consecutive months.

The method is effective when the predecessor firm size is larger than three employees, and at least 70 percent of employees 'move' to the new IRD number. Some extra restrictions are placed on the predecessors: the predecessor IRD number must cease to be used; and a majority of employees must move to the successor in the month immediately following the 'death' of the predecessor IRD number. In addition, certain industries are treated differently. In the agricultural industry, for example, it is common for groups of casual employees to move en masse between farms, so the predecessor firms in agriculture are restricted to a minimum of 20 employees. The aim is to minimise the number of false negative links (where we do not recognize that two employers in different periods are in fact the same) without introducing excessive numbers of false positive links between employers (where two records are linked but in fact refer to different employers). Clerical review of this process showed that the level of false positives was approximately 2 percent while the level of false negatives was at most 5 percent. Using this method, approximately 10 percent of firm deaths in any month are shown to be continuing businesses; this corresponds to up to 37 percent of job and worker flows due to spurious firm deaths. At the beginning of a tax year (April), the percentage of spurious births is seen to be as high as 17 percent. When comparing these results to those of the probabilistic matching described above, employer repair is carried out only for employers (these are the only firms filing an EMS form) and that employers are a subset of the universe of businesses on the LBF.

3.2.2 Planned integration of LBF information into the BF maintenance process.

The LEED project will move from being a development project to an ongoing source of official labour market statistics from late 2005. During 2006 and 2007 Statistics New Zealand plans to integrate the monthly outcomes of the probabilistic matching and the employee tracking information into the BF maintenance processes. It is expected that this will have a number of benefits:

- Improve the quality and timeliness of BF information related to creating longitudinal links. Existing geographic unit tracking relies on survey feed back with consequential lags resulting from the administrative system lags, survey contact, response time, and processing time. This can result in lags of many months and some units do not respond to the surveys at all. The current processes also do not attempt to cover small businesses. LEED will supply monthly information on both employee tracking and probabilistic matching across the entire BF population.

- Reduce compliance costs as there will be fewer Statistics New Zealand survey questions. The use of the administrative LEED information will reduce the need for questions on legal structure and ownership changes in frame update surveys.

- Another by-product of the LEED information will be a quality check at the broad geographic level of the geographic unit coverage of the BF. If there are a cluster of employees for an employer in a region where the BF records no geographic unit for an enterprise, is the BF missing a geographic unit? This will be the first administrative data source available in New Zealand that underpins the coverage of geographic units for enterprises that operate in more than one location. Currently, the quality of the geographic unit coverage of the BF for multi location businesses is largely dependent on survey feedback.

4. Small and Medium Sized Enterprise (SME) statistics

4.1 Statistics currently being produced

Central and Local Government in New Zealand focus strongly on policies and an environment that fosters, supports and grows SMEs within the economy. The SME sector is also actively represented by business organisations. This strong interest in the SMEs is reflected by requirements for official statistics that contribute to the planning and decision making processes related to understanding and growing the sector. Statistics produced help to quantify the number of SME businesses, their importance to the economy, and to monitor and understand changes over time.

The statistics office also requires a strong awareness of business compliance costs imposed by statistical surveys on SMEs. Statistics New Zealand has a strategic direction of maximising the use of administrative data in the production of these statistics.

The main source of SME statistics in New Zealand are Business Demographic Statistics derived annually from the Statistics New Zealand Business Frame (BF). Business demographic statistics provide an annual snapshot (as at February) of the structure and characteristics of New Zealand businesses. Statistics are available on a range of variables, including industry, region, institutional sector, business type and size (employment levels), and the degree of overseas ownership. Data can also be produced on the dynamics of businesses, by examining the entry and exit of firms (commonly referred to as births and deaths) between reference points.

The Business Demographic statistics are released each year in a formal news release and are also available from the Statistics New Zealand website (www.stats.govt.nz). This website includes a "table builder" facility that lets users construct their own cross tabulations within the restrictions of confidentiality rules. For example a user can construct a table by industry code for a specified region for an employee size group that gives a count of geographic (location) units and total number of employees meeting the specified conditions. These point in time aggregated statistics are used extensively for planning by local government and businesses.

The demographic statistics of a more longitudinal nature (showing entry and exit rates or growth rates for example) are typically used by central government and other organisations that are involved in economic and other policy formulation that requires in depth understanding of business dynamics. How do businesses grow, survive, employ more people etc. These organisations are typically sophisticated users of statistics, often the using the information within an academic framework.

4.2 Definition of SMEs in New Zealand

The generally accepted definition of SMEs in New Zealand is enterprises having 19 or less employees. This definition is used by the Ministry of Economic Development which annually use Statistics New Zealand business demographic information to produce reports on the structure and dynamics of SMEs in New Zealand.

Over 96 percent of enterprises in New Zealand employ 19 or fewer people. Eighty seven percent of enterprises employ 5 or fewer people. Sixty five percent of enterprises have no employees.

4.3 Challenges and developments related to SMEs statistics

As with any statistical information there will always be quality issues and user requirements to address, which have to be balanced against resource constraints and the availability of data. The challenges currently being faced by Statistics New Zealand with business demographic statistics include the following.

4.3.1 Users requirements for fine level SME statistics that exceed the accuracy of the available data.

As noted above, the predominant source of BF update information for small and medium businesses is tax data. This provides sound information for the enterprise unit, but is lower quality for the geographic unit physical location data and fine industry classification. The BF is designed to provide a quality frame for national level statistics. Users are provided with fine level regional business demographic statistics that fit within confidentiality release conditions, with a strong caveat on the weaknesses of the data.

Within compliance cost and resource constraints, options for improving the quality of the fine level regional SME data from the Business Demography information include two developments in 2005/06:
- The use of LEED data to identify if there are clusters of employees in a region with no geographic unit represented on the BF. This will help the quality at the broad regional level, not at the fine level required by some users.
- The test integration for two example regions of agricultural units on the BF with administrative land parcel ownership information. Users of agricultural statistics that use the BF have a strong interest in integrating this data with various land based data sources.

4.3.2 Employee size data

The BF employment size measure uses an administrative tax based definition of employees. Many working owners (e.g. sole traders, shareholders of companies) of businesses are not classified as paid employees in the tax system. This means that there are many small businesses on the BF with zero paid employees recorded, over 60 percent of the enterprises on the BF have 0 paid employees. Some of these will be working owner type businesses, others will be in industries that do not require labour (e.g. asset holding). But a significant number of the businesses have labour provided by working owner(s). This is an information gap for users of statistics on micro businesses.

The LEED project is developing an administrative source of data to identify the number of working owners for an enterprise. This is using annual tax return information that is beyond the scope of the employee salary and wage tax data used to calculate the number of paid employees. It is expected that this information will be available to the BF and business demographic statistics from 2007.

4.3.3 SME definitions

As noted in various OECD workshops and papers there are a range of definitions used to internationally to define SMEs. Currently the standard New Zealand definition is employee based at the enterprise unit level. But this has some weaknesses. For example the definition does not take into account:
- The activity of the business. Some businesses have a high capital base (eg asset owning) but have minimal employees. These businesses should not be defined as SMEs. The employment size measure could be supplemented with a maximum value of capital assets.
- The ownership structure of the business. An enterprise can be owned by another enterprise. These "group of enterprises" ownership links are held on the BF. Individually one of the enterprises may qualify as an SME but should the SME criteria be applied at the combined group level? Typically the members of the group have some access to the resources (e.g. capital, labour) of the wider group.

International comparability and transparency is important for all statistics. Frequently New Zealand statistics on SMEs are compared with other countries statistics on SMEs. To do this in a meaningful way is challenging, given:
- The different size of the economies. Many "large" New Zealand businesses fall within the SME definitions used in some countries.
- The coverage of the data sources (e.g. business registers) used as a source of the statistics when different countries SME statistics are compared. The BF coverage of business activity is relatively comprehensive and includes all businesses above a minimum level of tax activity while the business registers in some countries can have a more restricted scope for example only covering businesses with employees. The minimum business size included in the source of SME statistics will have a significant influence on the statistics.

4.3.4 Longitudinal analysis of SMEs

As described, in section 3 (above), administrative churn and ownership changes can distort or cloud longitudinal statistics on SMEs. Statistics for example to understand exit and entry rates. With the use of the LEED information and probabilistic matching the capability of the LBF to better support longitudinal statistics will be available from 2007.

4.3.5 Data availability

Existing business demographic statistics for SMEs have typically been restricted to variables available from the core BF (e.g. industry code, business type, employee numbers). Other Statistics New Zealand survey data (e.g. salary & wage data, financial data) is sparse for SMEs, as typically it is collected by sample surveys with a small sampling fraction for SMEs. Therefore any use for the production of statistics restricted to a population of SMEs will have high sample errors.

However with the increased availability of full coverage administrative data such as VAT data, employee wage data by enterprise and annual financial data the range of statistics on SMEs can be significantly extended. Statistics New Zealand plans to use the LBF integrated with the various full coverage administrative data sources to provide a much richer set of statistics on SMEs and other businesses.

References

Kelly N (2003) Repairing EMS Employer Longitudinal Links. Statistics New Zealand, www.stats.govt.nz.

Seyb A (2005) Statistics NZ Longitudinal Business Frame. Statistics New Zealand, www.stats.govt.nz.

Seyb A (2004) Repairing LBF Longitudinal Links. Statistics New Zealand, www.stats.govt.nz.

Eurostat Business Register manual

 http://forum.europa.eu.int/irc/dsis/bmethods/info/data/new/embs/registers/embs1_5.html

Ministry of Economic Development SMEs in NZ - Structure and Dynamics 2005
http://www.med.govt.nz/irdev/ind_dev/smes/2005/index.html

For more information please contact: Geoff Mead
geoff.mead@stats.govt.nz