# The Use of Administrative Data within an Integrated System of Economic Measurement - The Experience of Statistics New Zealand

## Introduction

Statistics New Zealand has regarded its economic collections as a comprehensive system for some time. As Statistics New Zealand has enjoyed increasing access to administrative data, we have been assessing how central a role administrative data can play in the overall system.

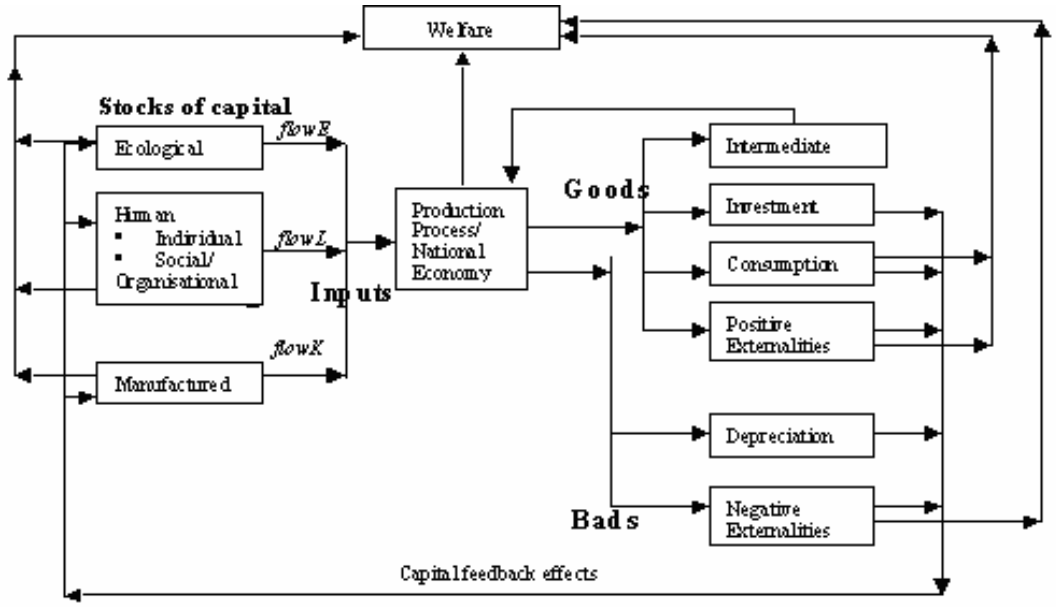
This paper explains the scope of the economic measurement system and how the diversity of user need increases the demand for the full range of measurement activity to be conducted coherently. This in turn creates the need for measurement infrastructure that facilitates the integration of multiple information sources and meets the need for comprehensive detailed information. The extent to which administrative data can be core to New Zealand's economic measurement discussion is discussed. The paper gives some specific examples of the way that administrative data has been used by Statistics New Zealand.

## Considerations in operating Statistics New Zealand's integrated economic measurement system

o *The diversity of user need*

The range of user needs that the economic measurement system has to meet has become more diverse over time. The System of National Accounts provides the organising structure for the measurement of market based economic activity, but an adequate set of economic data needs to embrace a wider set of social and environmental statistics that measure resource stocks and change over time.

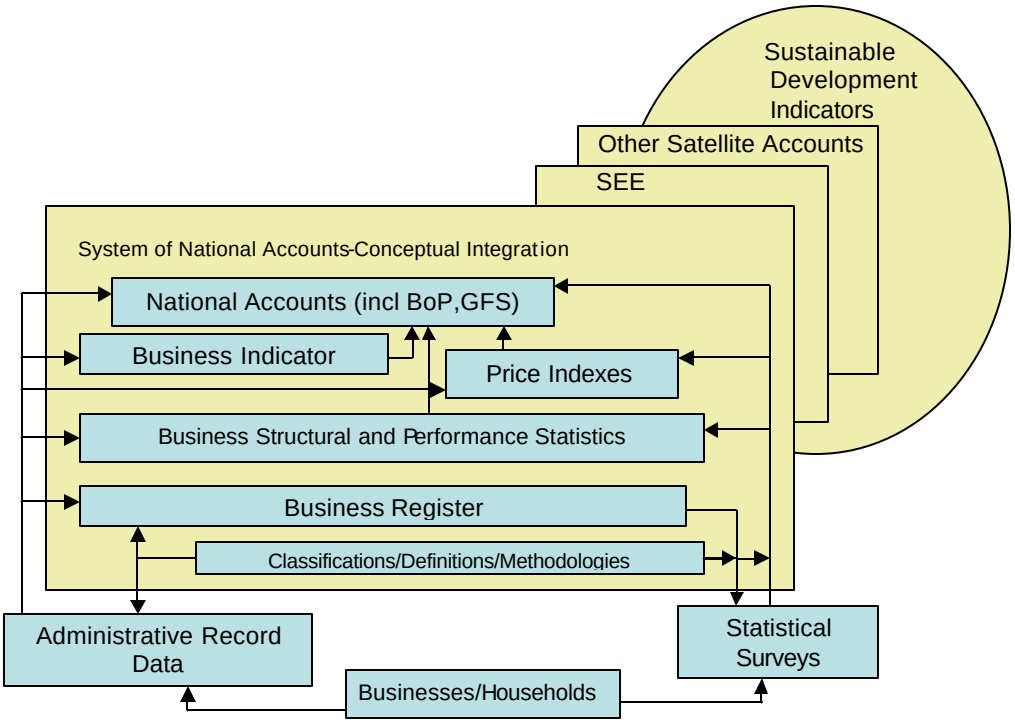
***Figure 1: Flows in the System of National Accounts***




## The system

The basic structure that Statistics New Zealand uses to meet these needs is common with most other countries. It is described in the following diagram.

***Figure 2: The New Zealand System of Economic Statistics***




User needs are being met by progressing with a four-pronged approach to collecting data to produce statistics about economic well-being:

- Structural analysis statistics including:
  - Annual financial collections and,
  - Periodic commodity collections
- Short-term measures collections.
- Indicator series collections.

o *Structural analysis statistics*

The strategy is to provide robust, annual data within 18 months of the reference year, which allows the identification and analysis of the structure and changes in the structure in the economy. The principle source of data is the Annual Enterprise Survey (AES).

The Commodity Collection and Business Price Index redevelopment programme currently underway will provide the basic breakdown information to supply index regimen weights and input-output ratios. These will be used in conjunction with AES to produce the national accounts within an input-output framework. Price indexes (required to do the input-output work in constant prices) will be kept up to date with periodic and cyclical collections relating to outlets and transactions to be priced.

Sustainability information is supported by the annual Agriculture Survey collecting agriculture production volumes data and land use data, periodic collections capturing the data required to measure natural national assets such as land, air and water.

o *Short-term measures*

This structural data is supplemented with economic data on a sub-annual basis that allows the main parts of the structural picture to be updated. The most significant areas of data contributing to the total picture should have measures produced on a monthly basis.

The existing sub-annual collections which were redesigned in 2001-2003 to collect sales and inventories include:

- Economic Survey of Manufacturing (quarterly) – (QMS) (also collects data on inputs and fixed capital additions).
- Quarterly Wholesale Trade Survey (WTS).
- Monthly Retail Trade Survey (RTS).

Data from a number of other diverse sources is currently used to estimate quarterly Gross Domestic Product (QGDP). The aim is to produce information, consistent with that produced for structural analysis, i.e. industry wide, which allows the earliest possible detection of turning points.

o *Indicator series*

Completing the data system are indicators. These are not necessarily financial in nature and are designed to aid the forecasting of yet-to-be-measured economic data. They should be available within one month of the reference period. Some of this data is also used in the production of QGDP.

The needs identified above place the following broad pressures on the measurement system;

- Data from different collections has to be able to used together. This varies from being able to reconcile aggregates within the national accounts framework to being able to do unit record level analysis of the relationship between business behaviour and business performance.
- Data collected for one purpose initially may have to be "recast" and used for another purpose. In practical terms this means that in addition to being combined with data from other sources, data will be used at greater level of disaggregation.

Both these demand create the pressure for data sources that are consistently reported across industries and have large enough sample sizes to sustain high levels of disaggregation. Administrative data meets this sort of requirement.

## Issues in the system

In looking to make improvements going forward there are a number of considerations that need to be integrated into development strategies.

o *Gaps in industry coverage*

As noted above, there are significant areas of the economy that are not explicitly covered by sub annual economic surveys. These include industries whose contribution to economic activity is significant and growing relative to those of the more traditional industries that are included in a data collections.

The foremost concern is the range of industries where estimates of contribution to QGDP are derived by extrapolating annual enterprise survey data, based on change

in labour inputs derived from the Quarterly Employment Survey (QES). In the Business Services industries, this is of particular concern.

o *Capturing structural economic changes*

The system needs to be responsive to changes business practices such as international globalisation, the move from a primary producing to service economy, and the uptake of new technology.

The introduction of the revised Australian and New Zealand Standard Industry Classification (ANZSIC06) will provide a number of challenges. The next round of changes will bring:

- The introduction of a new Information Industry division that will bring together activities involved in creating and distributing information based products. These activities are currently classified in manufacturing, communications and various service industries.
- The redefinition of the wholesale and retail industries along lines that reflect mode of operation rather than whether customers are households or businesses.

o *Growing demand from other users*

As well as the standard national accounting measures and the industry figures that have stand alone value within this system there has been a growing demand from other users in a number of domains. In annual financial statistics these include:

- Measuring the performance of sub-populations of interest e.g. tourism, Mäori and regional.
- Longitudinal micro-data analysis of financial data.
- Understanding the dynamics of business performance.
- Data integration with other Statistics New Zealand datasets such as Balance of Payments and surveys of business strategy and behaviour.
- An improved capacity to understand financial position, in particular the central bank of New Zealand wants to monitor financial stability of the non-financial corporate sector and from businesses for benchmarking purposes.

o *Pressure to reduce compliance costs for businesses*

The last review of the compliance cost of government activities for New Zealand businesses by the Business Cost Compliance Panel focussed 12 of its 108 recommendations on statistical issues.

The thrust of these recommendations centred around giving respondents the benefits of technology, more extensive communication with respondents, consultation over collection strategies, pursing strategies for reducing the number of requests for information and looking at ways to give respondents some of the benefits of having been part of data collection (to enable benchmarking their performance).

These recommendations added to an external environment in New Zealand where businesses expect government agencies to be actively reducing their demands for information and to streamline their procedures.

- *Need to develop survey systems to take advantage of new technologies that can increase efficiency and enhance analytical capability*

New technology can be of benefit not only to respondents, but also to statistics producers. Some examples in direct collections include the electronic capture and imaging of questionnaires, electronic spreadsheets that can automatically upload data into survey systems, and use of computer assisted technology such as CATI and XBRL. For administrative data, increased computing capacity provides the platform to hold and manipulate increasingly large amounts of data, and unique identifiers or "fuzzy" matching algorithms can be used to match large amounts of data prior to further manipulation and analysis.

- *General issues around the quality of estimates*

Examples of the type of issues that have been identified include;

In several of our quarterly economic surveys, many of the units in the panel sample had "outgrown" their strata. This made the survey estimates more variable and less robust as measures of temporal change.

The surveys in several industries subject to significant structural change over the life of their surveys (such as motor-vehicle assembly, meat processing, communications equipment manufacturing and boat building) required an increasingly high level of "special treatment interventions" to adjust individual responses. The amount of intervention required to maintain the survey estimates meant we were at risk of relying unacceptably on personal judgement rather than on the underlying survey design to provide robust statistics. It indicated that the variability of the estimates had increased significantly.

## The Centrality of Administrative Data in Statistics New Zealand's Measurement Strategy - steps towards a new measurement paradigm

A recurring theme surrounding these issues comes down to the capacity to organise the available information about the dynamics of the business population, store it on a central register and use it as a base to collect more detailed business information. Selective use of this information can help reduce the load which collecting data imposes on respondents, reduce processing costs, improve the coverage and reliability of statistics and improve the capacity to interpret data.

In particular:

- The measurement of economic variables has to be done as consistently as resources will allow across industries.
- Size and industry information has to be kept as up to date as possible, as "tired" survey designs can be a source of variability.
- We need to be able to enumerate population membership indicators for sub populations of interest. The main one of these is industry but the other populations, for example those within scope of Balance of Payment surveys, are also needed.
- We need to be able to track the ownership of significant assets (e.g. a large manufacturing plant) that have a physical location.

In broad terms the tax data available to Statistics New Zealand effectively allows Statistics New Zealand to:

- Maintain an accurate list of the population of active units.
- Maintain regularly updated size measures of employment, wages paid and turnover.
- Classify new businesses to industry, and to provide a range of triggers that indicate that a business' structure or industry might have changed so they can be prioritised for direct surveying.

Tax data can't adequately provide for new units and existing units:

- A structure of the main production activities of complex businesses.
- The tracking of the ownership of significant assets.
- Membership of sub-populations of interest.
- Changes in classifications e.g. change in industry.

There are other administrative sources that have the potential to meet some of these needs to varying degrees. The tax data can act as a "hub" to integrate data held at the Companies Office to update legal structures, workplace insurance data can update industry coding, and a range of other sources might be used to construct and maintain other sub-population identifiers.

The accumulated result of considering the contribution of administrative data this way is to develop a picture where:

- For the vast majority of small businesses with a simple organisation structure, administrative data is all that is needed to monitor their economic activity.
- The Statistics New Zealand register of businesses, the Business Frame, can provide the authoritative list of active businesses and their size and industry information. It accomplishes this mainly by being a hub for integrating administrative data from a range of sources and managing the concordances between administrative and statistical units.
- For businesses in the middle size range and across a range of industries administrative data provides a significant core of information comparable across the economy which can be supplemented by survey collection, much of which can be extracted from business information systems electronically.
- Sample surveys that are designed to complete the picture within the framework provided by administrative data, focus on larger businesses that don't have a simple one to one relationship between the administrative and statistical unit. The surveys can collect complex information particularly that not recorded in business' information systems.
- A range of analytical exercises that integrate administrative data sources and survey data sources and statistical infrastructure can produce new insights into emerging economic and social trends.

Administrative data can be used as a major source of frame maintenance information to provide the population dynamic and core economic information. In this picture the administrative data defines a context for survey collection and offers insights into some of the coverage and non-response errors in the surveys that are run. There are already many elements of this picture in place at Statistics New Zealand. As

survey systems are redeveloped, data from New Zealand's taxation authority, the Inland Revenue, is being used more extensively.

For example:

- Inland Revenue's Client Register provides timely demographic information about businesses.
- Goods and Services Tax (GST) data provides monthly, two-monthly or six-monthly sales and purchases for businesses registered for GST (Annual sales > $40,000 or voluntarily GST registered).
- Employer Monthly Schedule (EMS) data provides monthly counts of employees, earnings and PAYE for all employing businesses.
- Annual accounts information (IR10) provides annual financial performance and financial position data for around 65 percent of businesses on the Business Frame.

Some examples and lessons learned are given below. The remainder of this section offers some general observations about the type of considerations that have made this option an attractive one.

o *The characteristics of administrative data are well matched to the uses identified above.*

The key point here is that size measures, industry codes and some indicator variables can tolerate a certain level of error and still meet their purpose. The experience of operating surveys has highlighted that information being out of date is likely to be a bigger problem than measurement errors in this sort of classification information at the time of collection. The frequent updating is the key to quality.

The type of measurement principles that will underpin the approach of getting administrative data to underpin the measurement process are attached as an appendix. These general principles have been used in developing Statistics New Zealand's sub-annual collections.

o *Tax data has enormous value as a "measurement floor"*

As noted earlier the current sub-annual surveys do not cover all of the industries. The diagram below indicates those industries which are currently surveyed by Statistics New Zealand sub annually (in green shading) and those that are being considered for addition to the coverage (yellow shading) should these be justified.

The tax data from GST and EMS provide some measures, after some conceptual adjustment, of turnover and employment expenses. The picture illustrates how working on making the administrative data useable provides a core of information across the economy in a way that can be difficult to put in place by establishing new surveys.

***Figure 3: Sub-annual Industries – potential data sources***

| Coverage of SNZ Subannual Industry Collections | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Industry (ANZSIC Division) | | | | | | | | | | | | | | | | |
| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q |
| Agriculture, Forestry & Fishing | Mining | Manufacturing | Electricity, Gas & Water Supply | Construction | Wholesale Trade | Retail Trade | Accommodation, Cafes & Restaraunts | Transport & Storage | Communication Services | Finance & Insurance | Property & Business Services | Government Administration & Defence | Education | Health & Community Services | Cultural & Recreational Services | Personal & Other Services |
| | | QMS | | QBAS | WQS | RTS | RTS | | | | | QEAS | | | | RTS(pt) |
| | QES | QES | QES | QES | QES | QES | QES | QES | QES | QES | QES | QES | QES | QES | QES | QES |
| GST | GST | GST | GST | GST | GST | GST | GST | GST | GST | GST | GST | GST | GST | GST | GST | GST |
| EMS | EMS | EMS | EMS | EMS | EMS | EMS | EMS | EMS | EMS | EMS | EMS | EMS | EMS | EMS | EMS | EMS |

The generic approach for expanding industry coverage will use the Statistics New Zealand Business Frame as the source for the survey sample and population definitions. The variables collected will, as far as possible, follow a generic template which is consistent with the information requirements of key users, and will allow for inter-industry comparison and analysis on a consistent basis.

Administrative data sources, such as GST and EMS data available from the Inland Revenue, will be used where possible to model variables of interest as an alternative to directly surveying small to medium-sized respondents for these variables. This is consistent with Statistic New Zealand's compliance cost strategy to only survey businesses where administrative data cannot be used.

o *Administrative data can provide some of the economic variables*

Small businesses are simple units and administrative data combined from different sources can offer a relatively complete picture of the things that are worth knowing about a small business. Any limitations in this data are generally worth accepting in the context of the radical reductions in compliance cost that can be achieved. The amount of administrative data used, i.e. a $ value cut-off, will depend on the contribution that the selected small businesses make to the total contribution and how well the available data can be modelled to fit the variables of interest.

o *Answering questions beyond the scope of survey data: the potential of data integration*

The integration of administrative data into the framework of the Business Frame offers the scope to integrate administrative from a range of sources. The main feature in Statistics New Zealand's experience has been the scope to build longitudinal data bases and the capacity to combine personal and employment tax information with the full richness of the economic survey programme. These studies have the capacity to create new measures and do this at a greater level of detail than survey data could sustain.

This work will be the basis of Statistics New Zealand's efforts to understand dynamics in the economy. Again the key to the usefulness of this work is to ensure that it is well integrated with the Business Frame. Detailed discussion of this work is beyond the scope of this paper but the point is that it is the integration of administrative data with survey data that makes this work particularly valuable.

o *There is an attitudinal shift in making more extensive use of administrative data*

The greater use of administrative data has some common threads with capturing data directly from business' information systems electronically. Rather than relying on clarifying economic concepts for respondents and asking them to do their best to supply an answer, we move to the notion of taking data held by businesses in the form they store it and transform it into the concepts required by different outputs.

Statistics New Zealand has used a range of principles in guiding our work using administrative data. These are attached as Appendix A. They serve as a set of expectations about how we will do our survey design work. They aim to institutionalise the approach of using administrative data as the first option when fulfilling information needs.

## Examples of the Use of Administrative Data within this Framework

The following pieces of work are presented briefly to highlight some of the progress that has been made in taking full advantage of the characteristics of administrative data.

- The Business Frame work has highlighted the tremendous potential of using tax sources as both an underlying source of population information and as the hub for integrating administrative data with survey data, and for linking new administrative data sources into the measurement system.
- Some relevant experiences in making more use of tax data in sample design, in particular the capacity of tax data to provide comprehensive updates of quality across the economy, the capacity of tax data to provide data for a range of variables and the capacity of tax data to contribute to our understanding of non-sampling error.
- The use of administrative data in the 2002 Agricultural Census. This collection had been discontinued for a number of years and the issue of constructing a frame and the associated scope and coverage rules proved challenging.

**Case Study 1:**

## Integrating administrative data into the Statistics New Zealand Business Frame

The set of developments to make greater use of administrative data in the Business Frame processes were designed to exploit the timeliness and completeness of tax as a data source. The STEPS project (business Survey Tax Expanded Population Strategy) built a targeted set of direct surveying to sit on top of the complete tax information about the population.

The immediate outputs of the project were to create:

- A system that made employment and other tax measures available on the frame, and the process for updating these measures.
- Improved information about the "operating status" of businesses by making greater use of metadata from the tax system.
- Enhanced Business Frame processes using a range of tax data as "triggers" to detect significant changes in industry, size and structure.
- Enhanced flexibility of the Business Frame processes to make it easier to adapt the updating process, in particular enabling the integration of any administrative data that can be tied into the tax system.

The types of business gains that have been obtained are :

o *Reduction of lag in birthing new businesses*

The historical 12-16 week delay between business start up and availability for survey selection has been reduced to a 2-6 week delay. Previous studies of the effect of timeliness of appearance in survey populations suggested that this could produce biases of the order of the non-response bias.

o *Improved coverage of population for maintenance purposes*

The historical updating process only enabled partial updates of smaller businesses to be achieved over a 2-3 year period. The new facility provides updating information across the economy in the frequency provided by administrative systems.

o *Timeliness of the stratification variable*

The new employment stratification variable is updated monthly whereas the previous variable was updated annually for large units and considerably less frequently for smaller units. This creates the possibly of survey designs which re-classify businesses more frequently.

o *The direct surveying for frame updating has been significantly reduced*

The new systems have been designed so that as confidence in the use of administrative data grows, the amount of direct surveying can be progressively reduced. The pattern of the last few years has been to reduce the updating sample by about 10,000 questionnaires and halve the size of the monthly birthing survey.

o *Investigation into the use of other administrative data has been facilitated*

The systems have enabled investigation of the credibility of other administrative data sources. The enhanced facilities mean that alternative sources can be trialled and their scope for adding to the information about business population dynamics

assessed. An example is use of Companies Office data to provide information about changes in owners of companies.

o *A greater understanding of the population of tax units*

Analysts in economic statistical areas have benefited from understanding how the different tax forms can be used together and what these sources suggest about the level of and changes in economic activity. These analyses have helped put survey data into context in a way that wasn't previously possible.

Lessons Learnt

- The Meta data held by Inland Revenue has been very important in establishing the management of the business population.
- The Business Frame has become the main manager of the concordance of administrative reporting models and the statistical unit model. It is important that each new project that deals with administrative data doesn't re-invent these concordances and that they feedback what they have learnt about the quality of administrative data to the Business Frame.
- Investing in the capability of the Business Frame enables a range of improvements in surveys to be implemented progressively.

## Case Study 2 :

## Use of Tax Data in Survey Design

The sub-annual surveys have been redesigned along the general lines described above. Some general experiences can be outlined as:

o *Tax data has contributed to various improvements in sub-annual collections*

The availability of tax data has enabled a range of improvements to be made including:

- The use of GST and employment in bi-variate stratification to optimise sample selection.
- The use of GST and employment data to model variables for small to medium-sized units.
- The use of GST in incorporating regression imputation.

As a result of the use of this administrative data, the sample size within the QMS has been reduced by approximately 25 percent, while maintaining sample error requirements. This has been achieved notwithstanding the increase in the number of published industries from 10 to 17. The WTS and RTS designs also achieved similar results.

Research to date has focussed on the use of GST returns data. In the Manufacturing collection and Wholesale collections there has been successful modelling of a range of variables from administrative data. In the Retail Trade collection the GST data has been used to forecast values for smaller units. This builds on a history in our main annual collection in the use of tax data of different types, including the range of conceptual adjustments for which methodologies have been developed.

The main lessons have mirrored those of other agencies. There are useful gains to be made but the properties of the administrative data need to be understood and documented.

o *Potential for further gains in the annual collection*

As noted above, tax data has been used in the sub-annual survey redevelopments to achieve significant reductions in sample sizes. The Annual Enterprise Survey (AES) redesign in 1998 used Inland Revenue's accounts information for unincorporated business. This resulted in a compliance cost reduction of around 30 percent. Significant operational cost savings were also realised for AES as a result of the redevelopment of that survey.

During the next redesign of AES we are aiming aim to make further gains in reducing respondent burden by using tax data and Companies Office data for Incorporated Societies and Charities to cover small corporates and non-profit organisations. This approach has been successfully trialled in the commercial property industry.

We will investigate the possibility of developing a full coverage dataset of basic financial aggregates that includes respondent sourced data from the top 10,000 units by value, supplemented by tax data for the remainder. This dataset will support analysis for new statistics such as regional GDP and longitudinal studies. It will also provide the basic financial benchmark data that can be used to ask respondents for more detailed information such as commodity breakdowns and business performance measures.

We also want to develop a collection strategy to integrate all Statistics New Zealand requirements for annual financial data into a respondent focussed (rather than a survey focussed) data request including customised electronic questionnaires.

o *Size measures have added significantly to the robustness of designs*

Reselection is the practice of maintaining a repeating sample through reassignment to strata, but keeping the same strata boundaries and sampling fractions. This was introduced into the AES design in 1998 and gained more recognition when introduced for sub-annual financial surveys as a replacement for the panel sample approach i.e. selecting an initial sample and keeping these units in the same strata on an ongoing basis. The goal is to maintain samples over the life of the designs that are more representative of the contemporary population and also require less manual maintenance than the previous panel selected. This process has addressed the problem of sample degradation associated with units no longer accurately representing their originally assigned strata.

A related benefit has been the use of size measure in non-response follow-up. This has been a valuable tool in improving the operational efficiency of the survey process.

<u>Lessons learnt</u>

- This process has shown that the wide use of administrative data is meeting the sort of expectations outlined in the early part of this paper. Compliance cost reductions have been very well received and some significant sources of excess variability in our sub-annual collections have been directly addressed.
- Statistics New Zealand will continue to make more use of administrative data in its measurement system. The question that Statistics New Zealand will now be addressing is whether the best way ahead will be to keep working progressively in this way or to look at fundamentally rationalising our collections and redesign the information system more comprehensively. This is currently being reviewed.

- The type of changes that are underway will be reflected in Statistics New Zealand's system redevelopment processes. The greater use of data warehousing, meta data management and standardised business process will maximise the potential of these data sources.
- On a more technical level, the reselection and restratification processes have produced useful benefits but the full impacts of the way that this can be done and the way that sample selection across surveys can best be managed will still take some further development.

## Case Study 3:

## The 2002 Agriculture Census – Identifying an industry within this framework.

In 2002, Statistics New Zealand recommenced its cycle of measuring the Agricultural sector that was discontinued seven years before. First among many design issues was the challenge of re-creating an adequate sampling frame for the survey. The main issue was that there were farm industry sources that appeared as though they should provide appropriate coverage but the total numbers of farmers on these sources suggested that they were subject to significant undercoverage. This was confirmed by a number of studies undertaken by Statistics New Zealand.

The challenge was to consider why this survey couldn't be integrated into our Business Frame and identified initially by tax sources. The main difficulty arose in reconciling the notion of a farm as a parcel of land capable of agricultural production with the notion of a farm as an economic production unit that uses land as one among several factors of production. This was made particularly acute by the need to ask a series of questions around land use as a main focus of the collection. The anxiety was that the tax system has no need to collect industry data so there was some concern it would not be accurate.

The specifics of how this was resolved are beyond the scope of this paper but the main work occurred in constructing a set of operational "scope and coverage" rules to bridge the gap between the limitations of the frame and the conceptual requirements of the Agriculture Census. The tax based system did prove a success and it reinforced the point that it was better to have a frame that was too extensive and eliminate units that aren't eligible for the population than to accept undercoverage at the design stage.

A feature of how the work was completed was that different administrative sources were used to identify significant agricultural units and these were located in the tax system. This highlighted the value of having a comprehensive tax population as a reference list, the problems in using multiples sources (several matching processes had to take place) and the surprising limitations of the individual sources.

<u>Lessons learnt</u>

- The particular needs of the Agriculture collection required the standard frame processes to be combined with a set of carefully crafted operational rules to achieve a workable survey process.
- With the uncertainty around the establishment of an Agricultural frame it was beneficial to be building the frame in the context of a system that covered all industries. This meant that possible units could be found within the tax system and allocated to an industry.

## Concluding remarks

The demands of economic measurement will only continue to grow. They will build around standard measures and users will want to analyse standard aggregates against related measures like financial exposure and business behaviour. This analysis may be very intensive, and include designing the capacity for unit record analysis. Along side these demands, cost compliance pressures will not reduce.

Administrative data will be a major part of finding a way through these constraints. The precise way this can be done will depend on individual country's institutional arrangements. In New Zealand the high level of access to the tax system has created the opportunity for using administrative data to underpin the economic measurement system.

Our experiences have been encouraging. The capacity of administrative data to provide comprehensive, timely data of known quality has provided solutions to some ongoing quality issues. This work has mainly taken place in the context of improving the design of existing surveys. It has served to highlight how effective administrative data can be in fulfilling the need for core financial information.

The investment in the Business Frame, our success in integrating administrative data, and our plans to take fuller advantage of information management techniques have created a more extensive data infrastructure on which to design our future statistical programme.

It is vital that we manage the risks surrounding the potential risk of loss of access to administrative data. This is being dealt with by developing good relationships with administrative data suppliers, ensuring internal protocols and policies surrounding the use of administrative data are widely known, and regular consultation with the Privacy Commissioner and suppliers on data integration and micro-access issues.

We are now in a position where we are taking stock of our developments, looking at the next set of system and information gathering investments and evaluating the extent to which we will start to rationalise surveys and integrate the designs of existing surveys to each other. In any event it is clear that the survey collections will be smaller and will not need to meet all their traditional purposes, as our "measurement floor" of administrative data will have met this need.

# Appendix A:

## Data Collection Principles

*A. Parsimony Principle*

Only collect directly from respondents what must be collected after significant effort has been expended in avoiding additional collection activity.

Which variables in existing surveys are not published singly or in a derived quantity? Can it be proven that all published variables are used in a "valuable" fashion? If a variable is used only as a check, is that check implemented as a computerised rule and is that check of significant value? If not, drop that variable.

Which series are really used as measures of economic performance? Is some data collected for interpretation or validation purposes only, and could this be replaced with administrative data?

Can we use modelling techniques to supply additional detail or information?

*B. Best Supplier Principle*

Administrative data collected by an agency which has a statutory obligation to collect revenue or to administer government programs has a greater likelihood of being correct and complete in coverage than data obtained by a secondary agency.

What surveys re-gather data covered largely by administrative collections? Why aren't the administrative collections being used and development effort going into statistical estimation procedures to any overcome coverage or definitional problems? Can any samples be reduced in size because the administrative and direct survey approaches can be combined, at least to allow an early estimate feature? Given the budgetary realities, can delays inherent in switching to administrative collections be accepted?

*C. Reasonableness Principle*

Don't request data that cannot be supplied accurately and in a timely way from the respondent's accounting records.

The assumption might be made that if we request a non-standard accounting breakdown the respondent can comply without significant cost and will do so accurately. This is an un-businesslike attitude for us to adopt, and should be avoided unless there is no other way of establishing the degree of detail.

*D. Efficiency Principle*

A survey should aim to achieve the acceptable Total Error Measure (TEM = all bias and random errors) for the least cost.

TEM subsumes considerations of sampling and non-sampling error. Cost introduces budgetary constraints and the two components collectively measure efficiency. Consideration must be given to using administrative collections supplemented if necessary by surveys designed to provide sub setting detail on the grounds that the resultant TEM may be lower than that of a large, purpose-built survey.

*E. Robustness Principle*

Collection methodologies should recognise dynamism in collection framework definitions, population unit ownership and sampling structures in order that surveys can withstand the degradation of the original design circumstances without an undue rate of accuracy diminution.

Surveys designs should be efficient over their lifetimes, not just immediately following their inception. They should also be capable of responding to reasonable framework changes and accumulated commercial restructuring without minimum quality standards being jeopardised.

*F. Adaptability Principle*

Provide some ability to accommodate additional questions and new types of questions.

It should be possible to ask additional related business questions without undue delay. Also, the capability to request ad-hoc or cross-cutting questions should be considered in advance.

*G. Cognisance Principle*

Determine how each respondent is obtaining the figures supplied to Statistics New Zealand.

How do users incorporate the outputs from surveys in their own analyses?

We must face up to assessing quality of response AND planning for electronic data collection. We can only do this successfully if we know something of the respondent's information system (e.g. in-house and fully automated extracts, personal estimate, existing standard accounts or management reports, etc). The type of information system used by respondents might enable us to select samples by ease of data supply from the respondents' viewpoint.

If users use only top-level economic indicators and aggregates, reserving the more detailed aggregates and indicators to ancillary roles, then those latter statistics could be candidates for less costly estimation processes.