Tax Data in Official Statistics: Opportunities and Challenges

Dr Michael Slyuzberg
Inland Revenue, Wellington, New Zealand – michael.slyuzberg@ird.govt.nz

Abstract

The importance of the use of administrative data for the production of official statistics is recognized in Principle 5 of the United Nations Fundamental Principles of Official Statistics. A robust trend is observed for increasing the use of administrative data for statistical purposes. Statistical authorities use administrative data in two different ways – as a data source for published statistics, and in compilation of the sampling frame for statistical surveys of businesses. Sample data from administrative records and data bases are an important data source for the national accounts, regional economic accounts, and different economic and social research. In future, many statistical agencies would like to make more use of administrative data for statistical purposes in order to reduce compliance costs for business, particularly smaller businesses.

Data collected within the tax administrative system offers opportunities to complement existing sources of official statistics, to support wider uses of available data, or to minimise provider load within existing outputs. In the New Zealand context, use of tax data offers the opportunity to add value to the production of official statistics in a number of ways. Tax administrative data sources in New Zealand tend to be almost universal for some customer categories, that is, they may operate as an effective census. This can be done at comparatively low cost and with no additional response burden.

Using tax records for official statistics purposes, however, is only possible within strict legislative requirements aimed to protect both integrity of the tax system and taxpayer confidentiality – fundamental principles underpinning any tax agency.

The presentation covers the wide spectrum of applications of taxation data in the New Zealand official statistics. The examples include use of tax data in the Statistics New Zealand's Integrated Data Infrastructure, publishing tax statistics on the national and subnational level, release of confidentialised unit record files, and participation in the Open Data initiative. Key issues such as protecting privacy and confidentiality, managing data quality and collaboration with other parties are addressed.

Keywords: official statistics; tax data; confidentiality; response burden.

1. Introduction

In 1994, the United Nations Statistical Commission accepted the fundamental principles of official statistics (OS). Recently (January 2014) these principles were adopted by General Assembly. This acknowledgment at the highest political level emphasises the importance of OS for evidence based policy development and decision making.

The fifth fundamental principle specifically addresses sources of OS data. It states that "Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents." (see http://www.unece.org/stats/archive/docs.fp.e.html).

Recognition of administrative records as a legitimate (and valuable) source of official statistics opens doors for extensive use of tax data by national statistical offices as an alternative to traditional survey-based data sources. Tax data plays an increasingly important role in both economic and social statistics due to its obvious comparative advantages such as wider coverage, no additional provider burden and low cost.

New Zealand Inland Revenue (IR) considers information as a strategic corporate and governance asset. This approach is based on a few key principles, such as building the framework for effectively supporting and enabling IR's business goals and strategies; identifying business owners accountable for particular pieces of information, its content, integrity and quality; making all staff responsible for managing information as a corporate asset; simplifying and standardising metadata; describing information in consistent and understandable terms; easy accessibility of information for those who need it; and appropriate managing of information sharing to safeguard security and privacy.

This approach, in line with helping IR to make better decisions, develop more effective policies, better understand risks, and improve institutional knowledge, allows IR to improve the quality of tax data and successfully use it in official statistics.

There are multiple ways of using tax data in New Zealand official statistics:
- as raw data to produce statistical outcomes (for example, National Accounts, Balance of Payments, revenue statistics, income and expenditure statistics, productivity statistics, etc.),
- as unit level data supporting research and policy development,
- as statistical data directly disseminated by IR, and
- as confidentialised unit record files distributed by subscription via Statistics New Zealand.

2. Supply of tax data to Statistics New Zealand

Under current legislative requirements IR shares a substantial amount of the information it holds with Statistics NZ for the production of official statistics and research datasets. This includes (on a regular basis) client registrations, tax returns, employer monthly schedules, and additional data supplied from time to time on specific requests.

All these data are used by Statistics New Zealand to produce official statistics.

A Memorandum of Understanding (MOU) between IR and Statistics New Zealand established core principles of data supply between the two departments:

- Statistics NZ undertake all data integration exercises for statistical or research purposes and is the steward of the resultant datasets;

- Access to the integrated datasets will be managed in accordance with both parties' legislation, which practically means that:
  - access will be for research or statistical purposes only;
  - IR approval will be sought for all proposals to access the unit record data or datasets containing tax data; and
  - if there is any question over whether tax secrecy applies to any particular Information or data, including aggregate level Information, IR will be consulted and the IR view will be accepted as final.
- Any new data supply exercise should follow a comprehensive process with roles and responsibilities clearly prescribed to each side,
- Access to data sets that contain any data supplied by IR is provided by Statistics New Zealand subject to the clearly defined conditions after approval by IR on a case by case basis.

Access to anonymised unit level tax data is available for approved government and non-government researchers through the Integrated Data Infrastructure (IDI), a linked dataset created by Statistics New Zealand and made up of a series of datasets from different source agencies including Statistics New Zealand itself. These datasets include both longitudinal (such as the Longitudinal Business Database, a linked longitudinal dataset that covers a range of business-level data from surveys and various administrative sources) and cross-sectional components. All datasets have been integrated using both deterministic and probabilistic linking and confidentialised based on specific statistical techniques. The IDI is exclusively used for purposes such as policy evaluation, research analysis, and production of statistical outputs. Tax data is widely and successfully used within the IDI for research purposes. The last year topics included, for example, the impacts of minimum wage changes on the New Zealand labour market, firm behaviour and corporate tax reform in New Zealand, analysis of firm performance, productivity, innovation and skills, longitudinal productivity analysis, and exploratory analysis of tertiary education outcome indicators for graduates, to mention just a few.

In addition, we continue exploring other channels for secure dissemination of tax data. Recently IR jointly with Statistics NZ developed a confidentialised unit record file (CURF) containing a 5% sample of New Zealanders receiving salaries and wages in 2012, more than 110,000 records (see http://www.stats.govt.nz/tools_and_services/microdata-access/user-guides-for-curfs.aspx). Each record contains sex, age, region, industry, number of months a person received wages or salaries income, number of months a person received wage or salary income from two or more different employers, annual total taxable income and total income tax paid from this income. The sample design does not permit identification of particular people but allows various analyses on regional, demographic and industrial differences in the salaries and wages. For example, the CURF data has been used for research informing policy development related to wages and salaries taxpayers. The CURF is available to approved government and non-government researchers.

3. Producing publicly available tax data

IR maintains a strong position in the information-sharing domain, as it holds an extensive amount of information.

The New Zealand Government is committed to release high value public data to enable the private and community sectors to grow the economy, strengthen social and cultural fabric, and sustain the environment. In August 2011 the New Zealand Government approved the Declaration on Open and Transparent Government focussed on the effective re-use of public data. IR actively participates in this programme. We publish on our website an extensive set of downloadable tax statistics (more than 40 tables and graphs – see http://www.ird.govt.nz/aboutir/external-stats/) covering the period from 2001. Access to this data set is also available through the Government data portal (see

https://www.data.govt.nz/dataset/show/708). At the recent Official Statistics Users' Forum this dataset was mentioned as a convenient and useful research tool.

In particular, tax statistics includes:
- Registered customers by entity type and geographic areas,
- Active customers by entity type and geographic areas,
- Tax returns statistics,
- Number of tax agents and their clients by geographic area,
- Tax revenue statistics,
- Income distribution of individuals by income bands,
- Income distribution of salaries and wages individuals by income bands,
- Goods and Services Tax (GST) filers and net GST collected by turnover bands for businesses,
- Donations rebates (numbers and dollar values),
- Trust income data,
- Working For Families Tax Credit data (number of families receiving and average entitlements),
- High-level Child Support data,
- Filing of tax returns by channels (electronic vs. paper forms),
- Tax debt data by tax types and age of debt,
- Tax compliance data by tax type.

From 2013 we started publishing subnational tax statistics for selected variables on the Territorial Authorities level (see http://www.ird.govt.nz/aboutir/external-stats/regional-stats/regional-profiles-stocktake/regional-tax-statistics.html). New Zealand is subdivided into 73 territorial authorities, comprising of 15 cities and 58 districts. Tax statistics provides for each territorial authority detailed population data (structure, major age groups, registration for different tax types and social programmes) and information about tax debt and business turnover. This information is also available on the higher (regional councils) level. Release of these results required significant efforts related to proper geocoding of our data and cleansing exercises.

In close cooperation with Statistics New Zealand we publish Industry Benchmarks for Small and Medium Enterprises (see http://www.ird.govt.nz/industry-benchmarks/) – aggregated spreadsheets which include multiple financial ratios for selected industries. These benchmarks help business owners to work out if their business performance is similar to other businesses within the industry or differs from the industry benchmark range and why this may be. It allows the business owners to assess their performance. Statistics New Zealand has calculated the industry benchmarks using information on financial statements and tax returns provided by IR. All businesses supplying financial statements and tax returns were included in the calculation of the standard performance range, where the turnover for those businesses is between $60,000 and $10 million inclusive. Statistics New Zealand does not return any information about any specific business data to IR, nor anything that could be used to identify any individual.

In addition, IR produce multiple research reports containing rich information about New Zealand (see http://www.ird.govt.nz/aboutir/reports/research/). Some of these reports provide unique facts about important areas of New Zealand life – for example, a set of research reports about the impact of the Christchurch earthquakes on local businesses.

4. Improving data quality

Data quality is a vitally important characteristic of administrative data. It comprises multiple components such as relevance, timeliness, accuracy, comparability, accessibility and clarity. Low quality may result in unreliable or even misleading analytical outcomes, badly affecting decision support process and increasing reputational risk for data suppliers. This is why IR pays significant attention to data quality control. For example, the quality of externally published tax statistics is controlled on four levels.

Firstly, prior to raw data extraction data custodians review metadata (definitions) and update them if any changes have taken place since the time of the previous publication due to legislative or other reasons. If definitions have been changed, data custodians must amend existing data extraction queries, data sampling procedures (where applicable) and relevant pieces of published metadata. Information about previously used definitions is kept in the metadata dictionary available for users to avoid misleading comparison and support proper analysis of time series. Previously applied data extraction queries are kept in the query library for comparison and possible data recovery purposes.

Next, in the data extraction phase data custodians formally peer-review data extraction queries and, where applicable, data sampling procedures. Peer-reviewers feedback is documented.

Then, after data extraction all tables and graphs are thoroughly checked by a person responsible for the overall quality of external tax statistics publishing and signed off by a relevant manager.

Finally, after preparing data for publishing in web format, web pages are scrupulously checked by a publishing team in accordance to the current standards. This is followed by final sign-off at the executive level.

Published data is subject to a strict confidentiality control. Rounding, aggregation and suppression are used where necessary to prevent possible breach of privacy.

5. Challenges and tensions

All the above does not mean that the current state of using tax data in official statistics is ideal and free from problems. There is still much to improve.

As mentioned above, confidentiality and privacy protection is always our priority. This problem, however, is wider than just an "internal" issue of a particular government department. Our research shows that although public perceptions about reasonable data sharing are generally favourable they tend to change quickly in case of any adverse event – data leaking, human or system errors, etc. The change of perceptions relates not only to a particular department responsible for a privacy breach but rather affects the entire public sector. Even one serious breach of privacy may jeopardise a large amount of useful initiatives. We actively participate in ongoing discussions on minimising the risk of breaching privacy and confidentiality across a wider public sector including central and local government and academic institutions.

Another issue which needs serious attention is metadata management and standardisation of definitions. This is extremely important for the correct interpretation of analysed data, especially when analytical datasets include data from multiple sources. Discussions around metadata improvement continue for quite a while but there is still significant room for improvement.

Data availability is vitally important for all users of official statistics - analysts, researchers, policy makers and academics. Quite often they are surrounded by seas of data but do not know where to find a particular piece which they need. Knowledge about available datasets and well-developed search engines are needed as well as tools or platforms allowing remote and secure access to such data.

Another aspect of data availability relates to the data once used and becoming outdated. Users should have the most up to date data rather than once collected information which may become irrelevant and result in misleading analysis and poor decision making.

Despite significant efforts of a large amount of organisations, data quality is still an emerging problem for users of administrative data. This issue needs to be addressed to mitigate the risk of misleading decision makers by providing incorrect, irrelevant, outdated, unclear or not comparable data.

Quite often administrative data supplies a lot of information about particular departments and industries but not too much about our customers and communities, especially small groups with specific needs or issues. Further efforts are required to make administrative data more customer-centric and community-centric rather than "department-centric". A good example of such approach in New Zealand is the creation of the Integrated Data Infrastructure; however we consider it as a very important but still initial phase of developing a comprehensive customer-centric approach to official statistics data.

Last but not least, we need more people with advanced statistical and analytical skills able to use modern methods of quantitative and qualitative analysis, modelling, data mining and (not less important!) correctly interpret the outcomes and describe these outcome in plain language so it's understood by the target audiences. These skills may only be developed as a collective effort of all key official statistics stakeholders.

6. Conclusions

Tax data plays an important and continuously growing role in official statistics. The use of tax data as an alternative to traditional statistical data collection is driven by the intentions of national statistical offices to reduce production costs, lower response burden and extend population coverage.

In New Zealand tax data is widely used in the official statistics system. Applications of tax data include its supply to Statistics New Zealand and some other agencies producing official statistics, access to tax data through the Integrated Data Infrastructure managed by Statistics New Zealand, producing confidentialised unit record files, direct release of aggregated tax statistics and other data on the IR web site and through the Government data portal within the Open Government programme, and supporting relevant research and analytic projects.

Tax data released for use in official statistics sits within strict legislative requirements which protect the integrity of the tax system as well as the confidentiality and privacy of taxpayers.

Despite definite progress and a set of successful examples of using tax data in official statistics there is still significant room for development. The issues involve data and metadata quality, timeliness, availability, further improvement of confidentiality and privacy protection, making data more customer-centric and enhancement of more in-depth statistical and analytical capability to support decision making. These improvements require a coordinated effort from all key official statistics stakeholders.