

WORKSHOP 2: THE INTEGRATED DATA INFRASTRUCTURE - ENABLING CROSS-AGENCY RESEARCH AND EVALUATION

Statistics NZ

Abstract

Statistics New Zealand has created an Integrated Data Infrastructure (IDI) to make a wide range of data more accessible to users. The IDI contains longitudinal microdata about individuals, households, and firms. By bringing together many data from a number of sources, the IDI allows for longitudinal analysis across education, employment, migration, welfare, and business, with potential for further datasets to be added and linked. The IDI will help us to meet user demands for more or new information and to respond more efficiently to changes in existing data sources. It also represents an exciting opportunity to reuse and add value to existing administrative data.

The Enabling cross-agency research and evaluation workshop provided an overview of the IDI, including examples of research and evaluation currently being conducted, and a demonstration of the reach and versatility of the dataset. There were five separate presentations:

- *Building and using the IDI – Guido Stark, Kyle Uerata, and Brendan Mai, Statistics NZ*
- *Swimming in the IDI – Che Tibby, Statistics NZ*
- *Determining the outcomes of tertiary education: The contribution of the IDI – Zaneta Park, Ministry of Education*
- *Migrant outcomes and the economic impact of immigration – Paul Merwood, Ministry of Business, Innovation, and Employment*
- *Using the integrated data – Sarah Crichton, Ministry of Business, Innovation, and Employment Building and using the IDI*

Building and using the IDI

Guido Stark, Kyle Uerata, and Brendan Mai, Statistics NZ

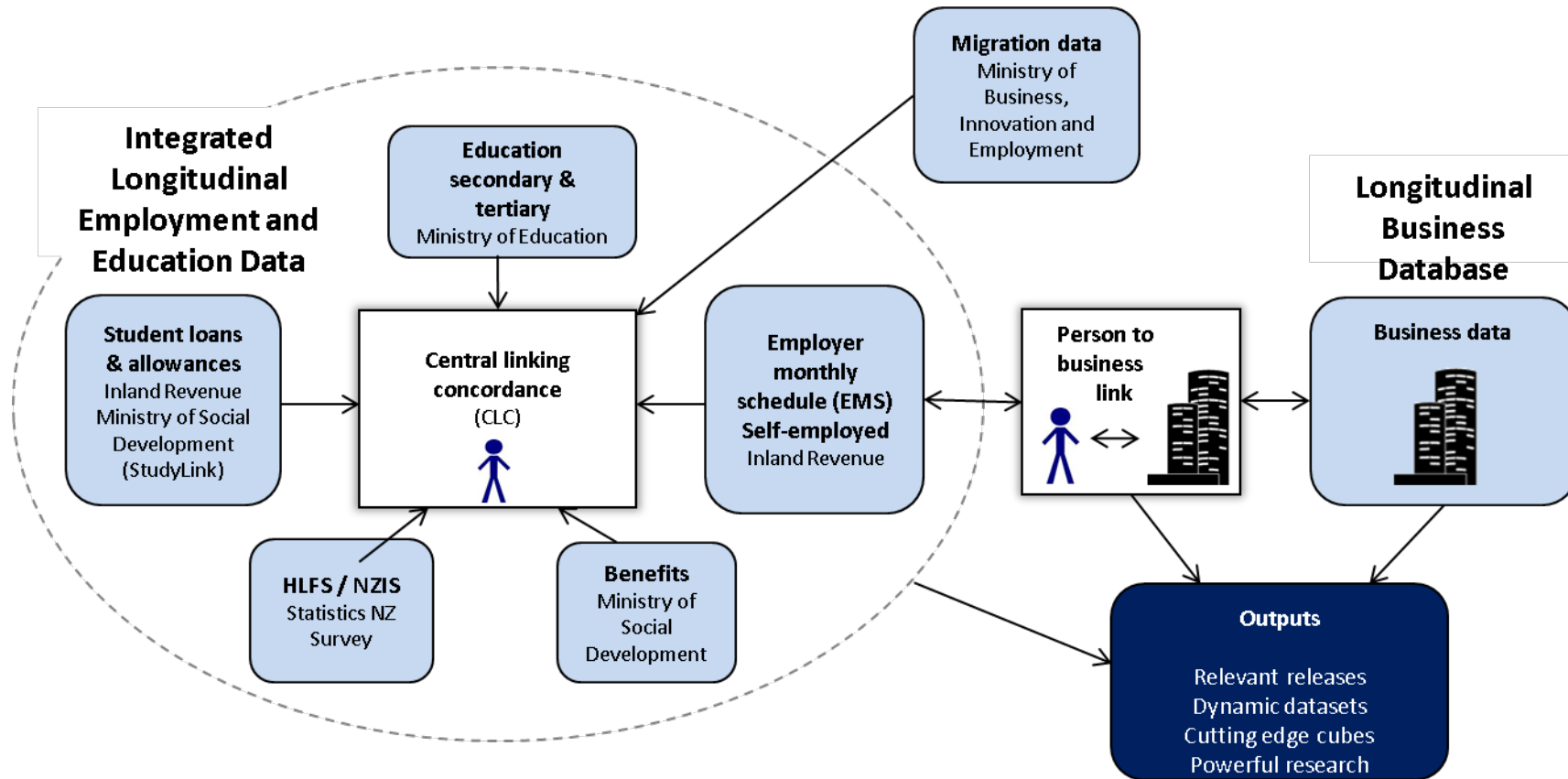
The IDI is based on a series of successful data-linking projects. Beginning in 2004 by linking data from Inland Revenue, Ministry of Social Development (MSD), and Ministry of Education, Statistics NZ created the Student Loans Integrated Dataset. This was followed in 2005 with the creation of the Linked Employer-Employee Database (LEED). Further data-linking projects include the Longitudinal Business Database, the Employment Outcomes of Tertiary Education feasibility study, the expansion of the Student Loans Integrated Dataset to include student allowance data, and expansion of LEED to include benefit data.

The IDI has been designed to ensure minimal manipulation and maximum utility, and to replace previous datasets. This data has been used by government researchers and has proven to be a valuable resource.

Key to the IDI is the development of the Central Linking Concordance (CLC). The CLC encrypts each set of data feeding into the IDI by allocating distinct unique identifiers. Once the data has been encrypted it is linked, as depicted in Figure 2.

The Integrated Data Infrastructure

Figure 1: The structure of the IDI



Specific links within the Integrated Data Infrastructure

Figure 2: Linking within the IDI – arrows indicate the databases that are linked directly.

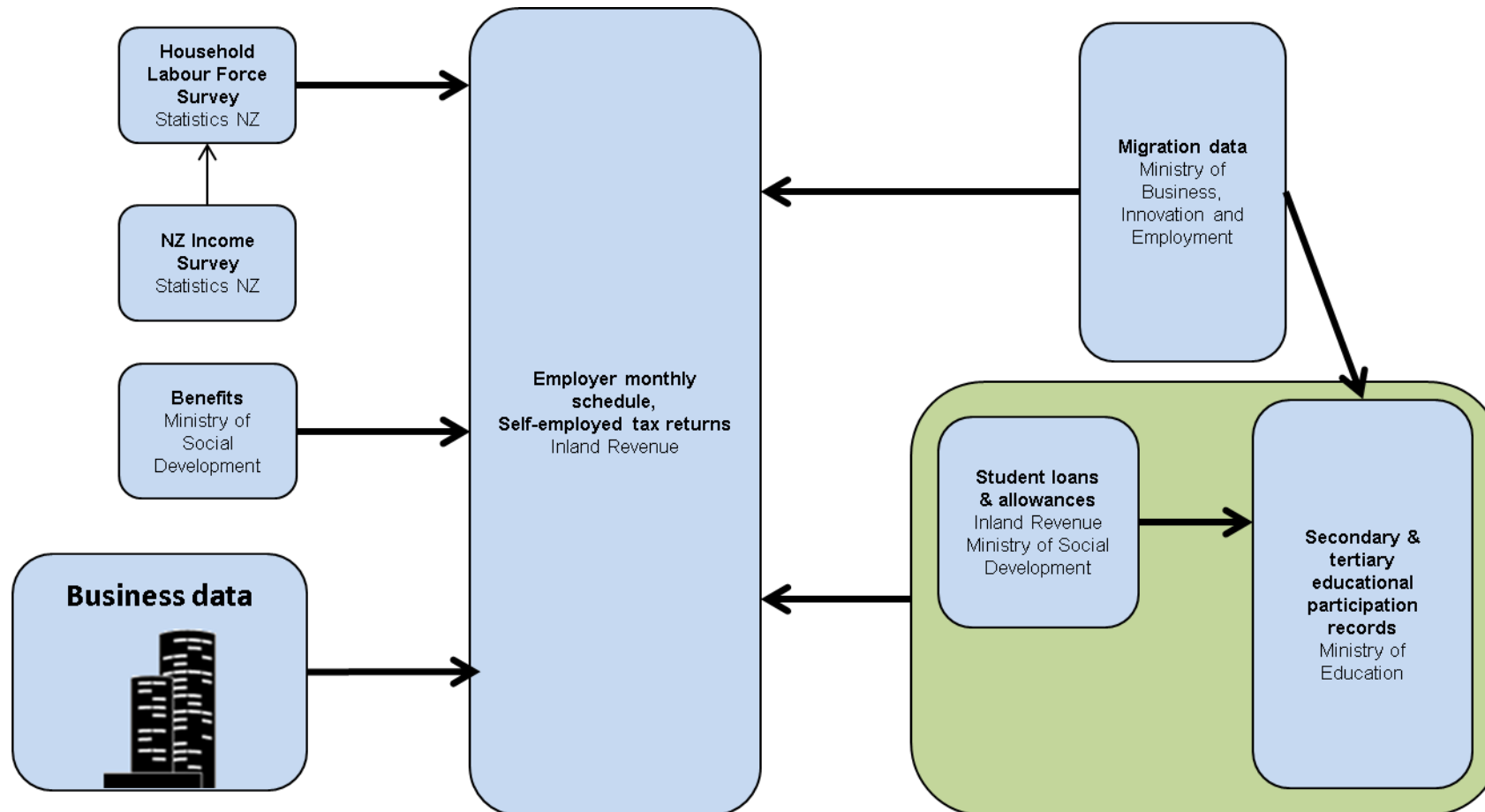


Figure 2 indicates the direction in which distinct sets of data within the IDI are linked. Data is linked via encrypted unique identifiers (such as IRD numbers or MSD social welfare numbers) and probabilistic matching (HLFS for example).

Accessing the IDI is relatively simple. The IDI is available to researchers who meet the specific requirements of the Statistics Act 1975 and the Tax Administration Act 1994. They access the data via the Statistics NZ Data Lab facility. When a researcher has a valid proposal and purpose, and is willing to conform to our rules concerning individual confidentiality and legal compliance, we welcome applications to use the IDI.

Swimming in the IDI

Che Tibby, Statistics NZ

The IDI captures a broad sweep of information with many possibilities. This presentation provided a glimpse into the lives of a group of 18-year-olds as they leave school and enter the workforce. Data accessed for this presentation included information on education, training, employment, migration, and benefit receipt.

With very large amounts of data it is possible to create multiple perspectives on the activities of both individuals and groups. To limit this data, Statistic NZ selected a group of around 64,000 individuals and tracked their progress over five years. The majority of these people were present in multiple databases, and were categorised according to their primary activity over the 60 months of the reference period.

The categories used were relatively simple, and mainly involved prioritising some activities over others. For example, a person undertaking paid employment and tertiary study in the same month was categorised as a 'student'. A person undertaking tertiary study while drawing a benefit (but **not** a student allowance) was categorised as a 'beneficiary'.

This presentation really only scratched the surface of the potential of the IDI. In aggregate, a number of effects and patterns were observable across the five-year period, including the global financial crisis, the movement of students and trainees into and out of education and training, and seasonal movement of people into and out of work.

Determining the outcomes of tertiary education: The contribution of the IDI

Zaneta Park, Ministry of Education

What do graduates do after they complete a qualification? And what do they earn? These are two key questions of interest to students, tertiary education institutions, policy makers and the general public. The IDI means that these questions can be answered more accurately than ever before.

The IDI makes it possible to better inform student choice about what and where they study, and to help providers

understand and improve student outcomes. Because the IDI is the only data source combining educational, border-crossing, benefit, and employment data at the unit record level, it is regarded as ideal for reporting. Previous work resulted in six separate papers published by the Ministry of Education. The IDI allows these papers to be built on by providing data on graduate destinations (what they do after completing qualifications) and employment outcomes (how likely they are to get a job, and how much they might earn).

The Ministry of Education plans to publish results and outcomes at one, two, and five years after graduation. The results will be presented by level and field of study, and will be published in an annual report. Individual reports for providers will also be published if the providers agree to publication.

Migrant outcomes and the economic impact of immigration

Paul Merwood, Ministry of Business, Innovation, and Employment

The IDI provides an opportunity to examine a wide range of migrant outcomes, as well as aspects of the economic impact of immigration, in ways that have previously not been possible. The Ministry of Business, Innovation, and Employment is asking a large number of policy questions concerning the economic impact of immigration, the effect of immigration on workplaces and workplace safety, and the impact of immigration on the labour market and labour productivity.

Migration research benefits from using the IDI because it allows groups to be followed over time, key economic indicators to be captured, and direct comparison between migrants and other New Zealanders. There are some limitations though. It is difficult to monitor very recent trends or policy changes, or to examine perceptions, ideas, and experiences (because the IDI is almost exclusively administrative records, with only limited survey data). Even IDI data does not make it possible to understand every aspect of people's working lives.

Specific research has included the ongoing monitoring of the Recognised Seasonal Employer scheme (including aspects such as return rates among workers, or employment duration and remuneration), and evidence to support a strategic review of the Working Holiday Scheme (such as who works, for how long, and what proportion of their holiday is spent working).

Future work may focus on ongoing reporting of migrant employment outcomes (earnings and sources of income), and the labour market impact of immigration (which industries are most reliant on migrants, and how this changes over time).

Using the integrated data

*Sarah Crichton, Ministry of Business,
Innovation, and Employment*

This presentation outlined the history of the IDI since initial efforts to establish LEED began in 2001. Between 2003 and 2006 the Department of Labour and Treasury demonstrated the value of LEED, producing a large number of publications. Following this exploration it was realised that expanding the range and type of data included would enable even greater value to be added by the data.

Since 2006, the addition of an increasing variety of data has enabled a much wider range of research. Early attempts to link in data were reasonably difficult, but this has become a lot easier and is expected to remain relatively simple.

The Ministry of Business, Innovation, and Employment is currently investigating several research possibilities. These include identifying the impact of further education and training on the earnings of working adults, the impact of active labour market programmes on jobseekers' labour market outcomes, and the characteristics of recent graduates who leave New Zealand for an extended period.