

# Statistics New Zealand's Linked Employer-Employee Data (LEED)

Presented to the HILDA Conference (July 2007)

Ian McGregor  
Project Manager (LEED)  
Statistics NZ

## Brief background to LEED data

The Linked Employer-Employee Data (LEED) project integrates longitudinal data on payments made to employees from which tax is deducted (sourced from Inland Revenue) with Statistics NZ Longitudinal Business Frame (LBF) data, to create new and enhanced statistics on labour market dynamics, including filled jobs, earnings, turnover and job creation and destruction. It also enables the production of 'person-level' statistics such as employment spells and tenure, earnings and income, and transitions between income sources and earning deciles. The census nature of LEED also allows levels of disaggregation – by industry, region, sector and firm size - not previously available. The starting year for longitudinal statistics produced from LEED is 1999.

The LEED project was driven by the need for better data to address a range of policy issues. LEED has the potential to improve the knowledge base for policy development in employment, business performance, firm creation and growth, employee turnover and many other areas.

The LEED initiative followed the successful development of similar datasets by a number of European and North American official statistical agencies.

## Development of LEED

Stage 1 of the LEED project involved Statistics NZ and the Department of Labour, in collaboration with Inland Revenue, investigating the feasibility of Statistics NZ integrating payroll data on businesses from Inland Revenue with other data on those same businesses held within Statistics NZ. This prototype integration demonstrated that LEED can provide improved information on business and employment performance and dynamics. Stage 1 was completed in September 2003.

Stage 2 of the Cross Departmental Research Pool funded element of the LEED project was completed 30 June 2004. This stage focused on:

- consolidating the methods and processes developed in Stage 1, including the development of a full prototype system and formal specification of the elements of a full production system
- undertaking specific labour market research projects using the prototype dataset
- identifying areas for future research
- release of technical and research papers on the Statistics NZ website.

## Moving LEED statistics into production

In 2005, the LEED project:

- made progress on the agreed research projects
- progressed the development of a number of statistical output
- migrated from prototype to full production version.

In 2006, the LEED project:

- embedded and refined regular official statistical outputs, both quarterly and annual
- produced four quarterly sets of official job-level statistics, each with a different focus, eg regional, industry, and sex/age
- released the prototype person-level outputs in June 2006 and the official annual person-level outputs in October 2006. The latter now incorporates information and statistics on self-employed persons, sourced from annual tax returns
- progressed the Ministry of Social Development/Statistics NZ feasibility project in terms of documentation and rules for exchange of data on beneficiaries
- published a range of research papers on topics such as benefit-to-work transitions, returning to work from injury, and worker-firm matching. The topics selected have been of policy relevance and have also helped us to understand the methods and data set. There is a rolling three-year research plan that has been agreed to by the LEED Sponsor Group
- commenced a methodology, output and dissemination review. The methodology and output review set out to look at six specific issues relating to LEED data that through refinement would lead to a better quality LEED dataset. One example of this is the birth-date imputation method which currently leads to a number of implausible birth dates. The dissemination review is considering alternative ways to publish LEED statistics other than the current Hot Off The Press method.

In 2007 LEED will:

- complete the output and methodology review and implement recommended improvements to the LEED database
- continue to update the quarterly and annual LEED Table Builder online, and publish analytical reports/Hot Off The Press as appropriate (dependent on outcome of dissemination review)
- release the quarterly publication with a focus on territorial authority statistics on 22 May 2007
- continue the development of an extensive research programme involving (at present), seven external researchers as well as on-going research by Statistics NZ analysts
- seek ways of publicising the LEED statistics more effectively
- attempt to identify and better meet the needs of users
- reach a mutual decision with the Ministry of Social Development (MSD) on whether or not the LEED/MSD data integration project will seek funding to go into production
- be involved in a data integration feasibility study with the Ministry of Education using education data on tertiary enrolments and completions.

### **Benefits of the LEED project**

Apart from the outputs mentioned above, the LEED project has had a number of other benefits.

- LEED has been a vehicle for building and consolidating good relationships with many of Statistics NZ's stakeholders. The collaborative model established early in the project with Inland Revenue and the Department of Labour, whereby researchers are able to be seconded to Statistics NZ to work on projects, while satisfying the relevant provisions of the Tax Act and Statistics Act, has assisted in our understanding of the data. Collaboration has also built a constituency for championing the project through the budget round and beyond and in dealing with privacy and confidentiality concerns.

- The LEED project has, along with other data integration projects in Statistics NZ, helped to develop capability within the organisation and secondees, although the project is still dependent on a small team of experienced analysts, developers, researchers and methodologists.
- The LEED project has demonstrated the value of the Statistics NZ ‘audience model’ by requiring us to carefully consider the different needs and communication preferences of our different audiences, and tailoring communications accordingly. For example, the dissemination strategy has included seminars and a session at the recent Labour Employment and Work (LEW) Conference; a series of seminars around the country for regional users; high-level Hot Off The Press releases that have been picked up and used by a range of regional and national media; and the inclusion of as much data as possible in Statistics NZ’s Table Builder facility, which allows users to specify more detailed ad hoc queries.
- LEED has, through its basis in administrative data, generated a number of spin-off benefits for Statistics NZ’s (and the wider Official Statistics System’s) statistical developments and business processes. For example, LEED data is being or will be used to establish the LBF and business demography statistics, to inform the development of regional GDP measures, and to confront GDP and employment data. LEED has the potential to complement or enhance existing statistics, which may in time lead to reduced respondent burden.

### **Issues encountered by the LEED team**

- The nature of the project has required a lengthy development cycle. While the lessons learned from LEED stand Statistics NZ in good stead for future projects of this type, time will always be required to explore and understand administrative datasets and their potential for producing official statistics, and this will be an iterative process.
- Meeting privacy and confidentiality requirements and concerns are obviously paramount, particularly when using tax data, and these considerations have placed a number of limitations on how the raw and integrated data is stored and used. Having said that, one of the contributors to the projects success has been Inland Revenue’s strong commitment to supporting the use of their data for broad public-good statistical purposes.
- While LEED has generated a range of new and more detailed statistics that fill gaps in the statistical framework, this has also brought with it the challenge of clearly articulating and communicating how the new statistical outputs relate to existing high profile administrative data (eg benefit statistics) and survey data (eg income, employment and wages statistics).

### **Privacy, security and confidentiality**

- Statistics NZ and Inland Revenue have an agreement that governs the transfer of tax data for statistical purposes.
- Inland Revenue data is encrypted prior to transmission and decrypted upon arrival into Statistics NZ.
- Unique identifiers (IRD numbers) are individually encrypted and names and addresses removed.
- Raw Inland Revenue data is stored on a separate server from the ‘cleaned data’ server.
- Both servers are separate from those used by rest of the Statistics NZ organisation.
- Access to data is strictly limited and controlled.
- No information is released by the LEED team that would allow for the identification of any individual or business.
- All data cells in any released tables comply with Statistics NZ confidentiality rules.
- Statistics NZ does not feed data back to Inland Revenue.

## Summary of LEED research

Currently there are seven part-time, seconded researchers working with LEED data. The Department of Labour has contracted researchers from Motu to work on the programme.

- Sylvia Dixon (DoL)
- Jason Timmins (DoL)
- Sarah Crichton (DoL)
- Dean Hyslop (the Treasury)
- Dave Mare (Motu)
- Steve Stillman (Motu)
- Moira Wilson (MSD)

LEED research reports published to date are listed here and are accessible from the Statistics NZ (LEED) homepage.

- Dixon [The Measurement of Labour Productivity using the Annual Enterprise Survey and LEED: Results of a data integration and data quality assessment.pdf](#) (May 2007) (441KB)
- Guzman-Posadas [An Exploratory Analysis of Aggregates and Individual Wage Earnings in LEED.pdf](#) (November 2006) (156KB)
- Moore [Survival Analysis of Transitions from Benefit to Work Using Administrative Data.pdf](#) (November 2006) (906KB)
- Stillman and Hyslop [Examining Benefit-to-Work Transitions Using Statistics New Zealand's Linked Employer-Employee Data.pdf](#) (November 2006) (482KB)
- Mare and Hyslop [Worker-Firm Heterogeneity and Matching: An analysis using worker and firm fixed effects estimated from LEED.pdf](#) (November 2006) (404KB)
- Dixon and Crichton [Successful Benefit-to-Work Transitions? The Longer-term Outcomes of People who Move from a Working-age Benefit to Employment.pdf](#) (March 2006) (352 KB)
- Crichton, Stillman and Hyslop [Returning to Work from Injury: Longitudinal Evidence on Employment and Earnings.pdf](#) (885 KB) (July 2005 updated August 2005)
- Hyslop, Stillman and Crichton [The Impact of Employment Experiences and Benefit-Spell Duration on Benefit-to-Work Transitions](#) (September 2004) (473 KB)
- Crichton, Stillman and Hyslop [Returning to Work from Injury: Longitudinal Evidence on Employment and Earnings.pdf](#) (September 2004) (881 KB)

At least one more paper will be completed and released by June 2007.

- The highest priorities for the remainder of 2007 are:
  - a. research into the employment patterns of older workers (particularly around the time of retirement), to support the Department of Labour's policy interest in older workers
  - b. an analysis of worker turnover and short-term employment patterns, to assess what insights LEED can contribute in support of the Department of Labour's policy interests in job matching, labour shortages, and casual and precarious employment
  - c. further progress with the 'Worker and firm heterogeneity and matching' stream of research, which is exploiting the unique features of LEED to investigate some fundamental aspects of how the labour market functions and adjusts.
- One new project has been added – on paid parental leave.

Future scoping and development work could usefully assess the potential for research into:

- the integration of person-level LEED data with firm performance data (IBULDD)
- the integration of LEED and MSD beneficiary data
- the integration of LEED and tertiary education data.

## **Current status of recent LEED data integration projects**

### *1. LEED and MSD feasibility project*

- Specified benefit data has been received from MSD.
- Documentation, such as privacy impact agreement, data extraction agreement and memorandum of understanding, has been completed.
- Inland Revenue has been fully informed throughout the process.
- Researcher from MSD is now writing up a small research project she has carried out using the data.
- Analysts from the LEED team are looking at matching of the data and also investigating potential outputs from the Statistics NZ perspective.
- A report summarising the success (or not) of the LEED/MSD feasibility study will be prepared by the end of August 2007.
- Should the final recommendation be to proceed with the creation of a LEED/benefit database for research purposes, a new funding bid will need to be put forward. The new database is unlikely to be available for a seconded researcher to use until 2008/09.

### *2. LEED and IBULDD*

- The IBULDD (Improved Business Understanding via Longitudinal Database Development) Project is an extension of the LEED Project to include wider business data.
- The first phase of feasibility (Jan 2006–March 2007) concentrated on bringing the data together to create the core IBULDD database. From this database, detailed statistical reports were produced analysing the overlap between the Annual Enterprise Survey (AES) and tax data.
- One aim was to test whether there was the potential to replace the AES questionnaires with tax data instead. The analysis identified the potential to avoid the need for approximately 4,000 AES questionnaires and thereby reduce compliance costs/burden on businesses.
- Following discussion at the international peer group review, the chosen method to fill gaps in the database was ‘donor imputation’.
- The later stages of phase 1 were spent producing examples of new official statistics.
- The second phase of feasibility (January–December 2007) concentrates on using the dataset more fully for research purposes, and identifying issues with the data/database.
- The IBULDD research programme is also designed to demonstrate the potential of data integration to support policy-orientated research, by using integrated data to address significant research questions in novel ways. It is hoped that by demonstrating the value of linking administrative and survey data, the cross-government research programme will build broader support for integration exercises in New Zealand.
- Aggregated LEED data has been combined with IBULDD to create an enriched firm-level research database.

### 3. LEED and Tertiary Education Data (EOTE)

- The EOTE (Employment Outcomes for Tertiary Education) Data Integration Project has gained funding from a successful Cross Departmental Research Pool bid as from July 2007.
- The stakeholders (Statistics NZ, MOE, DoL, and TEC) are currently working through the resourcing and planning issues associated with this study.
- The EOTE project will explore the feasibility of integrating data on tertiary enrolments and completions with LEED in order to:
  - provide official statistics on the employment outcomes for people who undertake tertiary education and training as well as new statistics on human capital, firm productivity and economic performance
  - assist government with policy decisions on investments in tertiary education.

#### **Other possible future enhancements of LEED**

- These include adding new administrative data to LEED. Other variables that have been suggested for possible addition to LEED include migration data, ACC data and a wider range of income support and social assistance data. Any addition of new data would require the establishment a new data integration project, including the required approval from suppliers and the Privacy Commissioner.
- Using LEED or other administrative data to complement, enhance or possibly replace existing survey data, with consequential reduction in respondent burden. For example, Statistics NZ is investigating the potential for LEED to partially meet wages data currently collected via the Quarterly Employment Survey. At this stage the lack of timeliness of LEED data is proving to be a large barrier in this regard. Administrative data sources such as payroll companies are another option.

#### **Issues with LEED data integration that need to be considered**

- The extent to which the intrinsic properties of LEED may limit its usefulness for extended statistical purposes. For example, LEED does not currently include some key variables (hours, occupation, detailed demographic characteristics), it is not very timely (it is currently published 14 months after the end of the reference period for job-level and 18 months after the end of the reference period for person-level), it only includes income from which tax is deducted, and it is based on the individual as the unit of entitlement (no family or household information). A number of these constraints can be addressed through the addition of new data sources, but other limitations will persist.
- The ongoing security and consistency of data supply (including its robustness to changes in policy). This is an issue with all administrative data supplies.
- The acceptability to suppliers, privacy advocates and the general public of maintaining and potentially expanding large integrated datasets that include highly sensitive information. There may be trade-offs between maintaining separate datasets (each for a specified purpose) and minimising supplier burden and increasing functionality through adding to LEED.
- The extent to which the proposals are likely to lead to new or improved official statistics. Statistics NZ carries out data integration, and Inland Revenue supplies tax data, for the purpose of producing official statistics, which leads to a number of constraints on, for example, access to microdata. While a research programme has been a very significant by-product of LEED, it is not the justification for the project, and any future enhancements will also need to lead to new or

improved official statistics. However, there are likely to be ways in which various aggregates can be made available for research purposes (eg via the IBULDD research programme).

## **Appendix 1 – Published Labour Market Outputs from LEED**

LEED outputs do not match the timeliness of existing labour market statistics and are unlikely to do so in the near future. Published labour market outputs from LEED include:

### ***1 Regular time series outputs (Hot Off The Press or in INFOS)***

Job-level series produced quarterly (one-way and two-way tables only):

- filled jobs per quarter, in terms of number and percentage change from previous year
- job flow<sup>1</sup> rates (job creation and destruction)
- worker flow<sup>2</sup> rates (employee accessions and separations)
- worker turnover rate
- mean/median earnings per job.

Person-level release, published annually, covering the following content:

- sources of income
- income transitions
- income spells and job tenure
- multiple job holders
- self employed.

### ***2 Detailed 20/20 tables***

Detailed 20/20 (Table Builder) tables are updated every quarter for job-level statistics, and annually for the person-level statistics. Much more detailed information is available from Table Builder than is possible to include in any one quarterly or annual statistical release.

#### **o Job-level tables**

For each of the variables mentioned above, separate tables are produced quarterly for the following dimensions:

##### *One-way (or one-dimension) tables*

[Table 1: LEED Measures by Age](#)

[Table 2: LEED Measures by Sex](#)

[Table 3: LEED Measures by Sector](#)

[Table 4: LEED Measures by Industry](#)

[Table 5: LEED Measures by Region](#)

[Table 6: LEED Measures by Firm Size](#)

##### *Two-way (or two-dimension) cross-tabulations*

[Table 7: LEED Measures by Age and Sex](#)

[Table 8: LEED Measures by Age and Sector](#)

[Table 9: LEED Measures by Age and Industry](#)

[Table 10: LEED Measures by Age and Region](#)

[Table 11: LEED Measures by Age and Firm size](#)

[Table 12: LEED Measures by Sex and Sector](#)

---

1 Job flows are the number of jobs created/destroyed by either new/ceased businesses or the expansion/contraction of employment by existing businesses.

2 Worker flows are the number of new employees added to (accessions) or removed from (separations) the payroll of an employer during a period.

[Table 13: LEED Measures by Sex and Industry](#)  
[Table 14: LEED Measures by Sex and Region](#)  
[Table 15: LEED Measures by Sex and Firm size](#)  
[Table 16: LEED Measures by Sector and Region](#)  
[Table 17: LEED Measures by Sector and Firm size](#)  
[Table 18: LEED Measures by Industry and Region](#)  
[Table 19: LEED Measures by Industry and Firm size](#)  
[Table 20: LEED Measures by Region and Firm size](#)

*Three-way (or three-dimension) cross-tabulations*

[Table 21: LEED Measures by Age, Sex and Sector](#)  
[Table 22: LEED Measures by Age, Sex and Industry](#)  
[Table 23: LEED Measures by Age, Sex and Region](#)  
[Table 24: LEED Measures by Age, Sex and Firm size](#)  
[Table 25: LEED Measures by Age, Sector and Region](#)  
[Table 26: LEED Measures by Age, Sector and Firm size](#)  
[Table 27: LEED Measures by Age, Industry and Region](#)  
[Table 28: LEED Measures by Age, Industry and Firm size](#)  
[Table 29: LEED Measures by Age, Region and Firm size](#)  
[Table 30: LEED Measures by Sex, Sector and Region](#)  
[Table 31: LEED Measures by Sex, Sector and Firm size](#)  
[Table 32: LEED Measures by Sex, Industry and Region](#)  
[Table 33: LEED Measures by Sex, Industry and Firm size](#)  
[Table 34: LEED Measures by Sex, Region and Firm size](#)  
[Table 35: LEED Measures by Sector, Region and Firm size](#)  
[Table 36: LEED Measures by Industry, Region and Firm size](#)

○ **Person-level tables**

For the person-level, separate tables are produced annually for the following outputs:

*Sources of income*

[Table 1.1: Source of Taxable Income](#)  
[Table 1.2: Median Annual Earnings](#)  
[Table 1.3: Source of Taxable Income by Proportion](#)  
[Table 1.4-01: Main Source of Taxable Income by Secondary Source of Taxable Income \(person counts\)](#)  
[Table 1.4-02: Main Source of Taxable Income by Secondary Source of Taxable Income \(median income\)](#)  
[Table 1.5: Main Earnings Source by Industry](#)

*Income transitions*

[Table 2.1: Flows Between Main Taxable Income Sources 1-year](#)  
[Table 2.2: Flows Between Main Taxable Income Sources 5-year](#)  
[Table 2.3: Flows Between Taxable Income Bands 1-year](#)  
[Table 2.4: Flows Between Taxable Income Bands 5-year](#)  
Tables 2.5: Flows Between Taxable Income Deciles 1-year  
[2001](#) | [2002](#) | [2003](#) | [2004](#) | [2005](#)  
Tables 2.6: Flows Between Taxable Income Deciles 5-year  
[Table 2.7: Flows Between Earnings Bands 1-year](#)  
[Table 2.8: Flows Between Earnings Bands 5-year](#)  
Tables 2.9: Flows Between Earnings Deciles 1-year  
[2001](#) | [2002](#) | [2003](#) | [2004](#) | [2005](#)  
[Tables 2.10: Flows Between Earnings Deciles 5-year](#)

### *Income source spells and job tenure*

[Table 3.1: Source of Taxable Income by Months Received](#)

[Table 3.2: Median Annual Earnings by Months Received](#)

[Table 3.3: Source of Taxable Income by Continuous Spell Length](#)

[Table 3.4: Length of Continuous Job Tenure \(industry\)](#)

[Table 3.5: Length of Continuous Job Tenure \(region\)](#)

### *Holders of Multiple Jobs*

[Table 4.1: Number of Jobs Held and Earnings](#)

[Table 4.2: Workers with Multiple Jobs by Industry](#)

### *Self-employment*

[Table 5.1: Main Earnings Source by Average Firm Size](#)

[Table 5.2: Wages and Salaries as Main Source of Taxable Income, and Self-employment as Secondary Source](#)

[Table 5.3: Self-employment as Main Source of Taxable Income, and Wages and Salaries as Secondary Source](#)

[Table 5.4: Firms With Employment](#)

[Table 5.5-01: Comparison of Employee Jobs in Firms With and Without Working Proprietors \(industry\)](#)

[Table 5.5-02: Comparison of Employee Jobs in Firms With and Without Working Proprietors \(region\)](#)