Australian Government

Productivity Commission

A rationale for developing a Linked Employer-Employee Dataset for policy research

Matthew Forbes
Patrick Jomini

The Productivity Commission

The Productivity Commission is the Australian Government's independent research and advisory body on a range of economic, social and environmental issues affecting the welfare of Australians. Its role, expressed most simply, is to help governments make better policies, in the long term interest of the Australian community.

The Commission's independence is underpinned by an Act of Parliament. Its processes and outputs are open to public scrutiny and are driven by concern for the wellbeing of the community as a whole.

Further information on the Productivity Commission can be obtained from the Commission's website (www.pc.gov.au).

A rationale for developing a Linked Employer-Employee Dataset for policy research

Key points

- Linked employer-employee datasets (LEEDs) have been used to study labour productivity, firm profitability, job creation, wage determination, and the effects of policies and business practices on employees and on firms.

    - LEEDs combine information about employees, their jobs and their employers in a consistent framework.

    - LEEDs allow researchers to separate effects associated with employees from those associated with employers.

    - Longitudinal LEEDs can be used to make causal inferences about the effects of policy change, study employment transitions and business growth.

- The ABS Foundation LEED is a prototype LEED based on personal income and business tax records for 2011-12.

    - The Foundation LEED contains few variables, which limits its use for policy analysis.

- A LEED could be a primary source of information about business and labour markets in Australia. A LEED that covered all businesses and individuals paying tax would support analyses at detailed industry and occupational levels.

- Producing a LEED for policy research, requires that the Foundation LEED:

    - make better use of existing source data by including more variables and reporting continuous values where possible

    - use existing source data to produce a longitudinal series

    - be accessible to researchers outside government.

- The lessons learned from producing the Foundation LEED will provide valuable guidance in producing a longitudinal Australian LEED.

- In the longer term, it would be desirable to incorporate information that is not currently collected as part of the taxation system in the LEED. This would ideally include:

    - employee hours (which might be collected with the shift to Single Touch Payroll)

    - the education levels of employees (which could be obtained by linking with Commonwealth higher education student records)

    - the method used to set pay (which might be collected through Single Touch Payroll)

    - a measure of capital stock (which would require further investigation).

The ABS Foundation Linked Employer-Employee Dataset (Foundation LEED) is an experimental dataset that seeks to remedy the absence of a representative Australian LEED. This research note provides an overview of how a LEED can be used to better inform policy.[1] It then describes the Foundation LEED in its current form, and provides suggestions to develop it into a tool for policy research.

The ATO and the ABS are to be congratulated for making the data available and pursuing the development of an Australian LEED. Collating information about both labour supply and demand in a single dataset will allow improved analysis of the factors affecting labour market outcomes. That said, more effort is required to transform the Foundation LEED into a dataset that can be used for policy analysis. The Commission would encourage the ABS to further develop the Foundation LEED as it could become a very important resource for understanding labour market and business dynamics, and designing better policy interventions over time.

1   What is a Linked Employer-Employee Dataset?

Matched data about employees and employers provide opportunities to analyse the effects of policies on labour productivity, job creation and the determinants of wage and employment outcomes. The linked information on the labour market and business performance is needed because 'the outcomes of interest in every one of these areas are jointly determined by workers' and employers' behaviour' (Hamermesh 1999, p. 38).

LEEDs are superior to datasets that provide separate information about employee behaviour and outcomes, and about the performance of businesses (box 1). This is because LEEDs allow researchers to identify the effects of policy changes on both sides of the labour market — on employees (labour supply) and employers (labour demand). Without linked data, analysis of labour market outcomes cannot account for the role of *both* employee and employer characteristics in explaining these outcomes (Jensen 2010, p. 209). For example, using the Housing, Income and Labour Dynamics of Australia survey to examine the determinants of wages does not account for the profitability of the employer, its investment and capital stock, or the degree of competition in the markets in which it operates. All of these factors affect wages and employment. Similarly, examining firm productivity using the Business Longitudinal Database does not account for the composition and characteristics of firms' workforces, which affect this outcome.

As noted by Hamermesh, the fundamental justification for linking employer and employee data stems from the fact that 'most labo[u]r market outcomes result from activities on both sides of the market' (1999, p. 26). A failure to consider firm effects when looking at the determinants of labour market outcomes can result in omitted variable bias, which can lead to inaccurate or misleading conclusions (Abowd and Kramarz 1999, p. 23). It is similarly

important to account for workforce characteristics (such as education and experience of the workers) when considering business outcomes.

Box 1   Types of Linked Employer-Employee Datasets

There are two main types of LEEDs:

- Cross-sectional LEEDS include a broad range of covariates at a point in time, including detailed information about workplace policies and practices, and human capital characteristics. Cross-sectional LEEDs are typically survey based.

- Longitudinal LEEDs track employees and firms over time, and are well-suited to analysing the relationship between the progress of the firm (labour productivity, growth in employment, firm survival) and that of its employees (wages, tenure). Longitudinal LEEDs are often based on administrative data sources. The time dimension can allow causal inferences about the effects of changes. It also allows some fundamental questions, such as whether productivity change can be attributed to a stable set of firms or to a process of creative destruction, to be examined.

The source of data for the LEED will affect the type of information included as well as the coverage of the dataset. A LEED derived from administrative data sources is likely to be limited by the purpose for which the data was originally collected, but may have a broad or even comprehensive coverage. An administrative LEED consisting of routinely collected data can be made into a longitudinal dataset more easily than a survey-based LEED. Conversely, a LEED derived from a survey is likely to have more limited coverage, but may include a richer source of information about workplace practices and outcomes. Survey-based LEEDs can contain specifically tailored questions that allow analysis of workplace practices (for example, details of workplace agreements) that are not be documented in administrative datasets.

It is possible to produce mixed LEEDs which combine administrative data with survey data, although data compiled from multiple sources can be difficult to use and interpret. This can allow longer-term analysis of the effects of changes in workplace practices (recorded in survey data) on firm and employee outcomes that are collected in longitudinal administrative data. LEEDs ideally distinguish between firms and workplaces, with the latter having a geographic location that allows location based policies to be analysed.

At the firm or activity/workplace level, LEEDs can include information about firm outputs, profitability, capital investment, revenue, workforce size and characteristics, location, industry, workplace policies and practices, adoption of new technologies, and workplace education and training. At the employee level, LEEDs can include variables describing workers' age, sex, labour market experience, education levels, occupation, wages earned, hours worked, job tenure and method of pay determination. LEEDs may also include information about the different jobs held by an individual.

Source: Bryson and Forth (2006).

Collating and linking information about employers and employees over time, in a longitudinal dataset, allows causal inferences to be made about factors that are associated with changes in either firm or employee outcomes (Bryson and Forth 2006, p. 2). As a longitudinal LEED tracks firms and employees over time, it can be used to identify whether a policy shift precedes or follows a change in employee or business outcomes. A longitudinal LEED also allows for more sophisticated analytical techniques where

unobserved variables, such as workers' cognitive skills, or a firm's entrepreneurial spirit can be taken into account.

A longitudinal LEED allows firm dynamics to be analysed. Changes at the firm level can be identified, including firm entry and exit, growth and decline, mergers and acquisitions, and changes in workplace practices. Employment dynamics include transitions between jobs, wage changes, geographical mobility, and movement in and out of the labour force. Where employee data is linked to transfer payment data, such as Centrelink payment records, it also becomes possible to analyse transitions between employment and unemployment (Hildreth and Pudney 1999, p. 3).

The use of administrative data is a critical factor in creating a successful LEED. It is difficult in a survey to ensure that both the employer and employee levels are *representative* of their respective populations. A non-representative sample means that any conclusions only apply to the sample and inferences about a policy change cannot be applied more broadly. This is likely to be a problem with LEEDs drawn from surveys where there are difficulties eliciting responses from businesses or their employees. Survey-based LEEDs, such as the 2014 Australian Workplace Relations Survey (FWC 2015), can suffer from low employer response rates.

To avoid sampling problems, LEEDs can be sourced from administrative data, which typically include records for an entire population. But it can be difficult to obtain appropriate economic variables from administrative sources (Hildreth and Pudney 1999, p. 3). For example, Chien and Mayer (2015) use a precursor to the ABS Foundation LEED to describe Australian firms. They note that the absence of key measures of labour inputs such as hours worked, educational attainment and experience, and the absence of information about capital stocks, limits their ability to examine labour productivity.

2   How does a LEED assist in policy research?

A LEED provides a unique perspective on the performance and outcomes of employers and employees. From a policy perspective, this includes an opportunity to identify and test what factors affect:

• labour market outcomes

• productivity at the firm, industry and aggregate levels.

Determinants of labour market outcomes

Labour market outcomes such as employment, tenure, worker displacement and wages are affected by a multitude of policies that influence both the behaviour of employees and the firms that employ them. Much of the research into the determinants of labour market outcomes focuses solely on the supply of labour — individual employees and their characteristics. Little attention is paid to labour demand. The role of employers in

determining labour market outcomes, other than possibly the industry in which they operate, is largely ignored. This is because the data is rarely available.

Ignoring the demand side is detrimental to analyses that seek to explain differences in labour market outcomes:

> [B]oth firms *and* workers play important roles in explaining observed differences in the earnings and productivity of individual workers; to ignore the effects of one would be to overstate the effects of the other. (Jensen 2010, p. 210)

By bringing together information about both labour supply and demand, linked data allows a more accurate analysis of factors affecting labour market outcomes. For example, this provides the opportunity to look at drivers of wage differences *within* firms, as well as across the labour market as a whole (Heinze and Wolf 2010). By taking into account firm factors, Heinze and Wolf (2010) find that the gender wage gaps differ across sectors in Germany. This type of evidence can be used to inform policies that seek to reduce pay inequality.

With a longitudinal LEED, it is possible to examine job creation and job destruction across the economy. Further, LEEDs can be used to examine costs associated with business closure, the results of which would be used to guide industry assistance policy (Hijzen, Upward and Wright 2006).

A LEED that covers all firms and employees can be used to explain how employment change occurs across industries, occupations or geography. For example, it could be used to understand how changes to the Automotive Transformation Scheme and other forms of assistance affect automotive manufacturing businesses or employees.

Similarly LEEDs can be used to analyse the effects of firm closure on displaced employees. For example, Margolis (2006) uses a French LEED to find that workers who experience unemployment following mergers or acquisitions often find employment relatively easily. This suggests that employment services do not need to target workers made redundant as a result of mergers and acquisitions.

In order to examine labour market transitions (from one job to another, or from low to higher paid employment, for example) it is necessary to follow employees over at least two periods to see how many remain employed and how many move to other firms or to other jobs. This is of particular interest for evaluating whether structural adjustment packages achieve their desired outcome. A longitudinal LEED that includes information about employee education, can be used to evaluate how education, employment and industry policy changes might affect employer and employee outcomes over time. For example, it could allow an analysis of the career progression of graduates, and an assessment of returns to investment in higher education.

Productivity outcomes

Productivity at the firm, industry and economy-wide levels is affected by the characteristics of both workers and employers. A LEED that provides information about business outcomes, and capital and labour inputs over time could be used to provide insights into how policy changes, combined with firm and employee characteristics, might relate to changes in productivity.

A LEED could, for example, assist in unpacking why leading firms can achieve high rates of productivity growth, while the overall average performance is underwhelming, as the OECD (2015) observed (figure 1). In this manner, Chien and Mayer (2015) use a precursor of the ABS Foundation LEED to describe characteristics of the most productive Australian firms.

Figure 1   **Productivity increases at the global productivity frontier[a]**

Change in labour productivity (percentage difference from 2001 values)
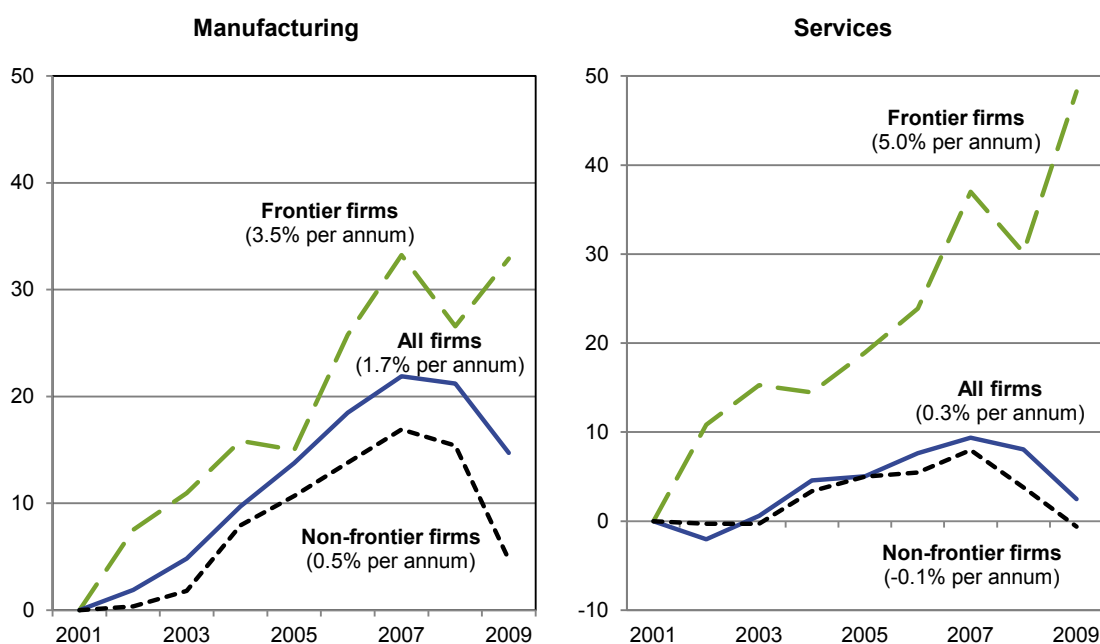
a The 100 firms with the highest average labour productivity in each 2-digit sector are considered frontier firms. All other firms are 'non-frontier' firms.

Data source: Figure 2.2, OECD (2015).

At the firm level, productivity is likely to be affected not only by the characteristics of the firm and the fortunes of the industry in which it operates, but also the characteristics of its employees (Guest and Stewart 2011). For example, causal inferences about the effects of human resource management practices on firm and workplace-level productivity and performance can be tested with longitudinal LEED data (Anderson et al. 2007).

The applications of LEEDs in this way are numerous. Where LEEDs incorporate information about workplace practices, technology use, industry regulation and research and development, they have been used to identify the association of these factors with productivity at the firm level (Andersson et al. 2008; Cottini, Kato and Westergaard-Nielsen 2011; Haltiwanger, Lane and Spletzer 1999; Mahlberg et al. 2013; Parrotta, Pozzoli and Pytlikova 2012; Turcotte and Rennison 2004).

Further, a LEED provides an opportunity to estimate productivity figures from the same data source — currently productivity estimates in the ABS' National Accounts are computed using estimates from separate surveys. If a LEED included detailed information on labour and capital inputs, it would be possible to calculate labour productivity and multifactor productivity estimates in greater detail — including at the firm level — than currently possible with the National Accounts. If the LEED were to become part of the National Accounts in the future, National Accounts estimates would become consistent with those derived from the LEED.

3   The ABS Foundation LEED

The ABS has recently constructed an experimental LEED that provides information on employees, their earnings, their jobs and the businesses that employ them. Referred to as the 'Foundation LEED', it is a prototype dataset built from 2011-12 personal income tax and business tax data from the Australian Taxation Office (ATO).[2] It was intended to demonstrate that constructing such an integrated dataset was both feasible and useful (ABS 2015a).

The Foundation LEED consists of separate employee, job and business files. Entries in the jobs file can be linked to both businesses and employees, with individual employees able to have multiple jobs, and be employed by multiple businesses. It includes information about 10.3 million individual employees who held 13.3 million jobs at about 683,000 businesses.[3]

Foundation LEED source data

The Foundation LEED is comprised of information from the Personal Income Tax (PIT) dataset, and the Expanded Analytical Business Longitudinal Database (EABLD) (ABS 2015a).

---

[2] The Foundation LEED discussed here is the product of the 'LEED Foundation projects'. The dataset referred to as the Foundation LEED in this paper is also referred to elsewhere as the 'Prototype LEED', the 'Experimental LEED', the 'LEED Foundation Projects Integrated Dataset' and the 'Experimental Statistics on Employee, Earnings and Jobs from Administrative Data, Australia 2011-12'.

[3] The Foundation LEED includes over 588,000 individuals who filed an income tax return, but could not be linked to a job or to an employer. It is possible that some of these represent owner-operators or workers on individual contracts. Further investigation is required to ascertain this, but it is likely that this LEED includes workers who are not employees.

The PIT dataset consists of person-level income tax data compiled for each individual who filed an income tax return with the ATO in 2011-12. Data items extracted from the PIT data for inclusion in the LEED include sex; age; occupation in main job; salary or wages; allowances, earnings, tips or director's fees; employer lump sum payments; attributed personal services income; employee share schemes data; fringe benefits; and employer superannuation contributions. The PIT data also includes an individual identifier that is used to link to other source data.

The EABLD includes financial information and business characteristics for all businesses active in the Australian economy in 2011-12 (ABS 2015b). It is based on the Australian Business Register, and includes data from three administrative ATO sources:

- *Business Activity Statements*, which report Goods and Services Tax obligations

- *Business Income Taxation* forms, which report taxable income or loss

- *Pay As You Go* (PAYG) data, which includes job-level information reported by employers about payments made to individual employees. In addition to gross payment amounts and information about dates of employment, the PAYG dataset includes employer and individual employee identifiers, which allows data from the EABLD and from the PIT to be linked.

Variables extracted from the EABLD for use in the Foundation LEED include information on the type of legal organisation, economic sector, industry, employment size and business turnover (ABS 2015a).[4] However, the EABLD is a very rich and detailed data source, and could be used to provide a more detailed picture of business activities.

Limitations of the Foundation LEED

Despite being a useful source of descriptive data, in its current form the Foundation LEED has two main limitations:

- As a cross-sectional dataset, the Foundation LEED cannot be used to examine trends over time, or to infer causal effects of policy change.

- The list of variables included in the Foundation is very limited (table 1).

A lot of important information, some of which is readily available from the administrative data sources, has not been included in the Foundation LEED.[5]

- The business file presents annual turnover and employment size variables as categorical variables. Converting continuous variables into categorical variables makes poor use of existing information, and limits their usefulness in quantitative analysis. The reasons

---

4 Despite being described in documentation as being extracted for the Foundation LEED, the type of legal organisation and economic sector variables were not in the Foundation LEED variable list.

5 It is recognised that accounting data, which is collected for tax purposes, may be very different from economic data. That said, the data provide a starting point for developing economically relevant variables.

for using categories are uncertain, but if for privacy reasons, then such concerns are unwarranted as there are a variety of other ways to ensure confidentiality without sacrificing the quality of the data.

- The business file excludes information about the legal status of the business, information required for the calculation of profits, expenditure data (including depreciation which could be used to estimate capital stock), research and development offsets, and tax paid. If there are concerns about the quality of these data, then including them for analysis would assist the ATO as researchers can utilise their skills to identify records where the data accuracy is uncertain. This means ensuring that researchers have a means to report back on the data issues they encounter. Such a feedback mechanism can assist both collectors and users of administrative data.

- The employee file excludes demographic information such as resident status, partner status, partner's income, superannuation contributions, different forms of income, information about tax offsets, deductions and expenses, and taxes paid. If the exclusion is for privacy or data quality reasons these issues should be addressed directly. If the aim is to keep the LEED to a manageable size, then attention is needed to identifying the most useful of the variables that are collected for administrative purposes. Usefulness should be assessed against the criteria of policy relevance — if a change in the variable is associated with different labour market or business behaviour, it should be included. Given the advances in storage, retrieval, and analytical methods, the number of variables should not need to be limited for technical reasons.

Table 1　　　**Foundation LEED variables[a]**

Employee file	Job file	Business file
Age (years)	Gross payment amount per job held during the reference period ($)	Annual business turnover ($) (categorical variable: $0-$49,999; $50,000-$199,999; $200,000-$1,999,999; $\geq$ 2,000,000; Not known)
Gender	Other jobs with which a job was held concurrently	Employment size (no.) (categories: 0-4; 5-19; 20-199; $\geq$ 200; Not known)
Total earnings from all jobs, in reference period ($)	Duration of job in reference period (weeks)	Industry (1 digit ANZSIC)
Number of jobs held in reference period (no.)		No business data available flag
Multiple jobs at the same time in reference period		
Occupation in main job (2-digit ANZSCO)[b]		
Geography (Statistical Area Level 4)		

[a] Files include identifiers to allow researchers to merge observations across the files. [b] An employee's main job is the job from which they reported the greatest earnings.

Sources: ABS cat. no. 6311.0.55.001.

How to make the Foundation LEED more policy relevant?

Discussions were held internally at the Productivity Commission and with external interested parties to consider what is necessary to improve the policy relevance of the Foundation LEED. Arguably the most important variable missing from the Foundation LEED is the number of hours worked by employees in their respective jobs. Without this information it isn't possible to quantify labour inputs, making it very difficult to use the Foundation LEED for labour productivity analysis or to work out any measure of hourly wages or labour income. The challenge with including hours worked is that this data is not currently collected by the ATO. There are also a number of other variables that are highly desirable to make the Foundation LEED useful for policy analysis.

The Commission has compiled a list of possible improvements (table 2). In compiling this list, the Commission has drawn on internal expertise and on staff in a number of policy departments (attachment A). The list identifies the variables that are available from the current data sources, and data sources for additional variables that meet the criteria for policy usefulness. There are, no doubt, other variables that could be accessed and would prove useful, but this list is a good place to start.

There was a consensus among those consulted that it is imperative that the Foundation LEED be extended longitudinally. Even if additional variables cannot be added, having longitudinal data will facilitate analysis of transitions and the labour market effects of some policy changes. A historical dataset going back a decade would be very valuable, but failing this it is imperative for policy usefulness that the dataset be built over time as a longitudinal product.

The use of actual rather than categorical values for variables was also viewed as important. The Department of Industry, Innovation and Science, the Department of Employment and the Fair Work Commission supported retaining values rather than making ranges available. Not only are continuous variable more informative, but many policies apply thresholds on income or activity for access to a policy treatment, so actual values are required to be able to consider behaviour of employees and firms around these thresholds.

Other improvements to the Foundation LEED require access to further data, either by linking to other data sources or by conducting surveys that use the Foundation LEED as a sampling frame. Those consulted differed in the prioritisation of what they were interested in seeing in a developed LEED. However, there was substantial agreement about the need for a measure of 'hours worked' to be included, in order to:

- be able to calculate hourly wages
- better quantify labour inputs in productivity estimates (as opposed to using employee head counts or wage bills).

It is possible that hours worked by employees will be routinely collected by the ATO in the future. The ATO is conducting a pilot program in 2017 to test the implementation of a Single Touch Payroll (STP) system (ATO 2015) (box 2). If STP were to report hours worked as part of the payroll data, this variable could be included in a future LEED.

Table 2 **Possible improvements to the Foundation LEED**

Change	Possible data source	Potential uses
Policy relevant and available from current sources		
Expand into longitudinal LEED	Existing ATO source data	Policy evaluation; employee mobility and transitions analysis; job creation and destruction analysis; industry-level time series analyses; effect of skill diversification on productivity.
Include more variables from ATO source data, for example:	Existing ATO source data	
- spousal income		Information about household income.
- superannuation contributions		Information about expected retirement incomes.
- company debt		Financing structure can relate to labour market and firm outcomes.
Policy relevant and available from other sources		
Educational attainment (employee variable)	HECS/FEE HELP data; or employee survey.	Estimate returns to human capital; productivity effects of education; gender wage differentials.
Hours worked (employee variable)	Survey (or single-touch payroll in future)	Calculation of hourly wages; labour productivity analysis; returns to education identification of low paid workers.
Method of determining pay (employee variable)	Survey (or single-touch payroll in future)	Estimate employment and income effects of changes to minimum wages and awards.
Transfer payments, job services data (employee variable)	Link LEED with DSS Research Evaluation Database[a]	Social policy and program evaluation (eg. How effective is job search assistance in helping people to achieve ongoing/ meaningful employment?).
Variable identifying new companies that arise through either mergers or acquisitions	Integrate with ASIC mergers and acquisitions data	Identify sources of business and employment growth.

a Personal income tax data is supplied to the ABS with a Scrambled Tax File Number (STFN), which uniquely identifies employees. The STFN is generated by the ATO before the PIT data is provided to the ABS. Deterministically linking the Foundation LEED data with other employee-level data would require that the ATO reproduce the STFN for these other data sources. Probabilistic linking of records is possible, but is likely to be less robust, and more expensive.

| Box 2 | The potential of Single Touch Payroll |

The Single Touch Payroll (STP) initiative is a proposed reform to the way businesses report their tax and superannuation obligations to the ATO. By integrating tax and superannuation reporting with standard business reporting software, STP is expected to substantially reduce duplication and record-keeping requirements faced by businesses. In turn, this is expected to ultimately lead to real-time reporting of payroll data:

> Under Single Touch Payroll, employers' accounting software will automatically report payroll information to the Australian Taxation Office (ATO) when employees are paid. (Billson 2015)

STP has tremendous potential as a source of employment-related data. As noted in the STP Regulation Impact Statement, the ABS has already recognised that STP could be used to identify turning points in the business cycle, and as a way of benchmarking employment survey data (ATO 2015, p. 30).

STP also offers opportunities to improve future iterations of an Australian LEED, given that it is closely linked to systems used to generate payslips for employees. Automated collection of payroll data raises the possibility that STP could be a valuable source of information about the hours individual employees work, whether they are casual or ongoing, and the industrial instruments by which their pay is set.

Other variables considered important on the employee side are as follows:

- A measure of highest educational attainment and skill would provide a measure of employee human capital. The HECS/FEE HELP data provides a source of this data for a substantial share of the workforce. A linked education-employment dataset has value in its own right as it supports assessment of the returns to public investments in tertiary education.

- Information about the method by which an individual's pay is determined would allow detailed examination of any association between the method of setting pay (such as, National Minimum Wage, Modern Award, Enterprise Agreement or Individual contract) and employee and firm outcomes. This would assist in the analysis of the employment effects of changes to the National Minimum Wage and to Modern Awards, as reflected in the Fair Work Commission's Annual Wage Review. Such information could assist in the design of future industrial relations policy. Information on the nature of wage setting could be integrated with ATO data collections in the future with the advent of STP (Billson 2015).

- Information about transfer payments (such as those included in the Research Evaluation Database[6]) would allow a better understanding of factors affecting transitions between welfare and employment. In particular, it would be useful to see how clients of job

---

6 The Research Evaluation Database (RED) is a longitudinal administrative database drawn from the Centrelink client database and maintained by the Department of Employment. RED includes transfer payment records and a range of personal details for all Centrelink income support payment recipients since 1998. Background Paper 5 of the Commission Research Paper into Housing Assistance and Employment in Australian contains further detail on RED (PC 2015).

search assistance services fare over time. This would allow the identification of circumstances when low paid employment serves as a 'stepping stone' to higher incomes, and when 'cycling' in and out of welfare payments is the result. A LEED would also allow the identification of the characteristics of firms where career progression is/is not occurring. It would be an ideal tool for evaluating the effectiveness of the current proposals to address youth unemployment through subsidised work placement schemes.

- Information about employer and employee superannuation contributions would facilitate the analysis of the effects of changes in superannuation policy on worker behaviour and expected retirement incomes.

On the employer side, more information about the dynamics of firm creation and destruction is considered important by DIIS. At present it is not possible to distinguish between new firms that are the result of mergers, acquisitions, spin-offs or franchises and those new firms that might challenge incumbents, or that are innovative start-ups. These distinctions would also assist in the identification of where new jobs are emerging in the economy.

Better use of the information already included in the EABLD would also improve the applications of the Foundation LEED. For example, the EABLD includes detailed information about investment. DIIS note however, that the EABLD does not include information about whether investment is financed by debt or equity. Further, Chien and Mayer (2015) note that the development of a measure of capital stock would be a useful addition.

In addition to developing the Foundation LEED by extending it longitudinally and expanding the number of variables included, it is important to make unit record data available to a broad set of researchers, including non-government researchers. Appropriate access requires that unit record data is made available in a secure manner that allows the datasets to be examined in a flexible way, as required by the researcher. Existing ABS data laboratory conditions where all outputs are vetted by ABS staff can facilitate this access. However, given the rapid progress of data analysis techniques in recent years, it is important that:

- the functionality of software not be limited beyond restrictions on networked communications.

- researchers be able to access open source software and user-written software libraries.

In allowing access to administrative data for research purposes other than those directly related to the administration of government programs or regulations, such as business taxes, agencies may face legislative barriers. However, the benefits of research that can look at policy questions in a more complete way, as is possible with a rich longitudinal LEED, go well beyond the specific program areas that collect the data. The data availability and use inquiry will be considering how best to address the legislative barriers to using administrative data for research purposes, as well as practical issues of extraction, compilation, curation and researcher access and use (PC 2016).

Attachment A

External consultations about improving linked employer-employee data

Name	Organisation
Dr. Luke Hendrickson	Department of Industry, Innovation and Science
Dr. Greg Connolly	Department of Employment
Workplace and Economic Research Section	Fair Work Commission
Prof. Mark Wooden	Melbourne Institute of Applied Economic and Social Research
Prof. Elizabeth Webster	Centre for Transformative Innovation, Swinburne University
Vera Holenstein	Macroeconomic Conditions Division, The Treasury
Marty Robinson	Social Policy Division, The Treasury
Matt Maloney	Tax Analysis Division, The Treasury

References

Abowd, J.M. and Kramarz, F. 1999, 'The analysis of labor markets using matched employer-employee data', *Handbook of labor economics*, vol. 3, pp. 2629–2710.

ABS (Australian Bureau of Statistics) 2015a, *Construction of Experimental Statistics on Employee Earnings and Jobs from Administrative Data, Australia, 2011-12*, ABS Cat. no 6311.0.

—— 2015b, *Construction of the Expanded Analytical Business Longitudinal Database, 2001-02 to 2012-13*, ABS Cat. no 8171.0.0.

Anderson, J.O., Lin, H.-S., Treagust, D.F., Ross, S.P. and Yore, L.D. 2007, 'Using Large-scale Assessment Datasets for Research in Science and Mathematics Education: Programme for International Student Assessment (PISA)', *International Journal of Science and Mathematics Education*, vol. 5, no. 4, pp. 591–614.

Andersson, F., Brown, C., Campbell, B., Chiang, H. and Park, Y. 2008, 'The effect of HRM practices and R&D investment on worker productivity', *The Analysis of Firms and Employees: Quantitative and Qualitative Approaches*, University of Chicago Press, pp. 19–43.

ATO (Australian Taxation Office) 2015, *Regulation Impact Statement: Single Touch Payroll*, October, Canberra.

Billson, The Hon. B. (Minister for Small Business) 2015, *Government moves to get Single Touch Payroll right*, Media release, http://bfb.ministers.treasury.gov.au/media-release/050-2015/ (accessed 26 May 2016).

Bryson, A. and Forth, J. 2006, 'The policy relevance of linked employer-employee data', in Bryson, A., Forth, J. and Barber, C. (eds), *Making Linked Employer-Employee Data Relevant to Policy*, DTI Occasional Paper, Department of Trade and Industry, London, UK, pp. 1–8.

Chien, C. H. and Mayer, A. 2015, *Use of a Prototype Linked Employer-Employee Database to Describe Characteristics of Productive Firms*, Research Paper, 14 May, ABS Cat. no. 1351.0.55.055.

Cottini, E., Kato, T. and Westergaard-Nielsen, N. 2011, 'Adverse workplace conditions, high-involvement work practices and labor turnover: Evidence from Danish linked employer–employee data', *Labour Economics*, vol. 18, no. 6, pp. 872–880.

FWC (Fair Work Commission) 2015, *Australian Workplace Relations Study (AWRS) - Provisional AWRS fieldwork report*, Fair Work Commission.

Guest, R. and Stewart, H. 2011, 'The Age Dispersion of Workers and Firm Productivity: A Survey Approach', *Australian Journal of Labour Economics*, vol. 14, no. 1, p. 59.

Haltiwanger, J.C., Lane, J.I. and Spletzer, J.R. 1999, 'Productivity differences across employers: The roles of employer size, age, and human capital', *The American Economic Review*, vol. 89, no. 2, pp. 94–98.

Hamermesh, D.S. 1999, 'LEEping into the future of labor economics: the research potential of linking employer and employee data', *Labour Economics*, vol. 6, no. 1, pp. 25–41.

Heinze, A. and Wolf, E. 2010, 'The intra-firm gender wage gap: a new view on wage differentials based on linked employer–employee data', *Journal of Population Economics*, vol. 23, no. 3, pp. 851–879.

Hijzen, A., Upward, R. and Wright, P. 2006, 'Using Linked Employer-Employee Data to Estimate the Earnings Costs of Business Closure in the UK', *The Policy Relevance of Linked Employer-Employee Data*, DTI Occasional Paper, pp. pp. 134–153.

Hildreth, A. and Pudney, S. 1999, 'Econometric issues in the analysis of linked worker-employer surveys', *Micro-and Macrodata of Firms*, Springer, pp. 509–540.

Jensen, P.H. 2010, 'Exploring the uses of matched employer–employee datasets', *Australian Economic Review*, vol. 43, no. 2, pp. 209–216.

Mahlberg, B., Freund, I., Cuaresma, J.C. and Prskawetz, A. 2013, 'Ageing, productivity and wages in Austria', *Labour economics*, vol. 22, pp. 5–15.

Margolis, D.N. 2006, 'Compensation policy, human resource management practices and takeovers', in Bryson, A., Forth, J. and Barber, C. (eds), *Making Linked Employer-Employee Data Relevant to Policy*, DTI Occasional Paper, Department of Trade and Industry, London, UK, pp. 28–44.

OECD 2015, *The Future of Productivity*, Organisation for Economic Co-operation and Development, Paris.

Parrotta, P., Pozzoli, D. and Pytlikova, M. 2012, 'Does Labor Diversity Affect Firm Productivity?', No. 6973, IZA Discussion Papers, Institute for the Study of Labor (IZA).

PC (Productivity Commission) 2015, *Housing Assistance and Employment in Australia*, Commission Research Paper, Canberra.

——— 2016, *Data Availability and Use*, Issues Paper, April, Productivity Commission, Canberra.

Turcotte, J. and Rennison, L.W. 2004, *Productivity and wages: Measuring the effect of human capital and technology use from linked employer-employee data*, Working Paper 2004-01, Department of Finance (Canada).