ECONOMIC STATISTICS CENTRE OF EXCELLENCE

A COLLABORATION WITH Office for National Statistics

A Roadmap for Developing a New LEED Infrastructure for the UK

John Forth, Alex Bryson and Christina Palmou

ESCoE Technical Report TR-28

June 2025

ISSN 2515-4664

TECHNICAL REPORT

A Roadmap for Developing a New LEED Infrastructure for the UK
John Forth, Alex Bryson and Christina Palmou
ESCoE Technical Report No. 28
June 2025

Abstract

Linked employer-employee data (LEED) are a valuable component of any country's data infrastructure. However, existing sources of LEED in the UK have significant limitations. This report argues that the UK now has a strategic opportunity to build a new LEED infrastructure around the employment and earnings data from the HM Revenue and Customs (HMRC) Pay-As-You-Earn Real Time Information (PAYE RTI) system. These data are comprehensive in coverage, longitudinal, high-frequency and linkable to other datasets. The report proposes that the new LEED infrastructure should be based around a linked employer-employee spine, which is augmented with data matched in from administrative databases and surveys. The resulting infrastructure should be periodically updated, curated and made available to the research community under controlled conditions. The report sets out a roadmap for the development of this new LEED infrastructure, identifies relevant stakeholders, and outlines ways in which the data could inform policy-making on issues such as business dynamism, productivity, wage growth, inactivity and job quality.

Keywords: Linked employer-employee data; data linkage; productivity; job quality

JEL classification: C80; C81; J2

John Forth, Bayes Business School, City St Georges, University of London
john.forth@citystgeorges.ac.uk

Published by:
Economic Statistics Centre of Excellence
King's College London
Strand
London
WC2R 2LS
United Kingdom
www.escoe.ac.uk

ECONOMIC STATISTICS CENTRE OF EXCELLENCE

A COLLABORATION WITH Office for National Statistics

A Roadmap for Developing a New Linked Employer-Employee Data Infrastructure

John Forth, Alex Bryson and Christina Palmou

ESCoE Technical Report No. 28

June 2025

A Roadmap for Developing a New LEED Infrastructure for the UK

John Forth, Alex Bryson and Christina Palmou

ESCoE Technical Report No. 28

June 2025

Abstract

Linked employer-employee data (LEED) are a valuable component of any country's data infrastructure. However, existing sources of LEED in the UK have significant limitations. This report argues that the UK now has a strategic opportunity to build a new LEED infrastructure around the employment and earnings data from the HM Revenue and Customs (HMRC) Pay-As-You-Earn Real Time Information (PAYE RTI) system. These data are comprehensive in coverage, longitudinal, high-frequency and linkable to other datasets. The report proposes that the new LEED infrastructure should be based around a linked employer-employee spine, which is augmented with data matched in from administrative databases and surveys. The resulting infrastructure should be periodically updated, curated and made available to the research community under controlled conditions. The report sets out a roadmap for the development of this new LEED infrastructure, identifies relevant stakeholders, and outlines ways in which the data could inform policy-making on issues such as business dynamism, productivity, wage growth, inactivity and job quality.

Keywords: Linked employer-employee data; data linkage; productivity; job quality

JEL classification: C80; C81; J2

John Forth, Bayes Business School, City St Georges, University of London

john.forth@citystgeorges.ac.uk

Affiliations

Alex Bryson is Professor of Quantitative Social Science at University College London, UK. He is a Research Associate of the Economic Statistics Centre of Excellence (ESCoE), UK, a Research Fellow at the Institute of Labor Economics (IZA), Germany, and a Faculty Fellow at Rutgers University, USA. He is also a member of the research team for the ADR UK-funded Wage and Employment Dynamics (WED) project.

John Forth is Professor of Human Resource Management at Bayes Business School, City St George's, University of London. He is a Research Associate of the Economic Statistics Centre of Excellence (ESCoE), UK, and a Research Fellow at the Institute of Labor Economics (IZA), Germany, and a member of the research team for the ADR UK-funded Wage and Employment Dynamics (WED) project.

Christina Palmou is a co-author in her personal capacity. She is a Lead Economist in the Economic Microdata Transformation team at the Office for National Statistics and an ESCoE PhD student at King's College London. From April 2025 - April 2026 she is on secondment to the Productivity, Innovation and Entrepreneurship division at the OECD.

Acknowledgements and disclaimer

This research has been funded by the Office for National Statistics as part of the research programme of the Economic Statistics Centre of Excellence (ESCoE). We thank David Freeman, Sanjiv Mahajan and Tom Wickersham at the Office for National Statistics (ONS) for comments on an earlier draft of the report. We also gratefully acknowledge comments from participants at a workshop on 'Developing New Linked Employer-Employee Data for the UK', held in London on 12th September 2024, and a meeting of the ONS Economic Experts Working Group in January 2025. We are grateful to Emma Gordon (ADR UK) and Mark Brewin (HMRC) for providing valuable insights into the legal framework for data sharing and data access. A number of other individuals have also provided helpful insights on the potential value of a new linked employer-employee data infrastructure for the UK. We cannot list them all, but particular thanks go to Maria Balgova, James Black, Alexander Hijzen, Josh Martin, Stuart McIntyre and Mairi Spowage.

Some of the analytical work for this report was undertaken in the Office for National Statistics Secure Research Service using data from the ONS Business Structure Database (ONS, 2024a). This does not imply the endorsement of the ONS.

Contents

Executive Summary

Introduction

Linked employer-employee data (LEED) are a valuable component of any country's data infrastructure, because they provide a more comprehensive understanding of the dynamics of the economy and labour market than studying organisations and households in isolation.

The shortcomings of existing sources of LEED mean that the UK currently has a significant gap in its statistical and research infrastructure. It implies that economic and social policy making is either based on UK LEED with significant limitations, or it is based on insights developed from LEED in different economies and labour markets, namely the US and Europe.

A strategic opportunity

This report argues that the UK now has a strategic opportunity to build a new LEED infrastructure around the employment and earnings data from the HM Revenue and Customs (HMRC) Pay-As-You-Earn Real Time Information (PAYE RTI) system. These data have the potential to overcome many of the limitations of existing sources of LEED in the UK:

- They are **comprehensive in coverage**, covering almost all employee jobs in the UK.
- They are **longitudinal**, with unique identifiers enabling employee records to be grouped within employing units and linked across time.
- They are **high-frequency** and **regularly updated**, by virtue of the real-time nature of the supply of payslip information by employers to HMRC
- They are **linkable** to other datasets, either using the unique identifiers in the data or using the new demographic and business indices developed by the Office for National Statistics (ONS).

We propose that the new LEED infrastructure should have the following components and features:

- It should be based around a **linked employer-employee spine**, containing unique and time-consistent identity linkages which allow employees and employers to be tracked together over time
- The spine should be **augmented** with additional employee and employer characteristics, matched in from administrative databases and surveys
- The data should be **periodically updated and curated** in a transparent and replicable way
- The data should be made available to researchers as a **public good** under controlled conditions.

Policy relevance

A new LEED infrastructure with these features could provide timely and authoritative information on a wide range of policy-relevant issues. The report outlines a number of illustrative use cases, covering issues of direct relevance to government, such as business dynamism, productivity, wage growth, inactivity and job quality. A new LEED infrastructure would also provide the basis for a swathe of economic and social science research that would enhance our understanding of the economy and labour market.

Curating a core infrastructure

In principle, there are already opportunities for accredited researchers to use the PAYE RTI data in the HMRC Datalab and Integrated Data Service (IDS), conditional on obtaining permissions from relevant data owners. However, we are proposing that the ONS, together with HMRC, the Economic Statistics Centre of Excellence (ESCoE) and Administrative Data Research UK (ADR UK), take the lead in both constructing, documenting and maintaining a canonical LEED infrastructure, built around PAYE RTI, in order to avoid a proliferation of LEEDs which are constructed in a less-than-transparent fashion, potentially contain erroneous linkages, are difficult to replicate, or in other ways fall short of best practice when designing a key data asset. We therefore emphasize the importance of **data curation**.

The development of this LEED infrastructure is technically feasible, given the support of relevant stakeholders, and the potential value is significant.

A roadmap for new LEED

The report sets out a three-stage roadmap for the development of a new LEED infrastructure:

- **Stage 1:** Utilise the PAYE RTI data to develop a **LEED spine** which will provide unique and time-consistent identity linkages for employees and employers, allowing workers and firms to be tracked over time
- **Stage 2:** Develop **linkages** from the LEED spine to a range of existing administrative and survey data sources, so that additional data on employers and employees can be added to the spine
- **Stage 3:** Use the spine as a **sampling frame** for new employer-employee surveys, so that gaps in our existing measurement of the employer-employee relationship can be addressed.

To move forward, we propose that ESCoE, ONS and HMRC should collaborate over the next twelve months to develop a ***prototype*** for a new LEED infrastructure, consisting of a LEED spine linked to a small number of additional datasets. In doing so, ONS would build on its unique position as the UK's National Statistical Institute (NSI) with unrivalled capabilities to create research-ready datasets through data linkage. It would also build on its experience with the PAYE RTI data and its existing relationships with HMRC, as owner of the PAYE RTI data. ESCoE would contribute academic expertise in the creation and analysis of LEED. Additional collaborative partnerships should also be sought with others in government and academia who have experience of PAYE RTI.

The goal, however, is to go beyond a prototype to create a sustainable LEED infrastructure that exists for many years to come, akin to the Longitudinal Employer-Household Dynamics (LEHD) infrastructure in the United States or the Linked Employer-Employee Data infrastructure (LIAB) in Germany. Therefore, attention must also be paid to developing the prototype into a long-term data asset, owned and updated by the NSI, linkable to other survey and administrative datasets and fit to address immediate and emerging policy needs.

To achieve this goal, government departments will have an important role in supporting data linkage. ADR UK should also have a key role in supporting these linkage projects under its next phase of funding, beginning in April 2026. Because the LEED spine will cover all workers paying tax in the UK, we envisage government and ESRC may use it as a sampling frame for new surveys. Additional surveys will be necessary, both for economic measurement and policy purposes, to enrich the administrative data with data items that more comprehensively capture the employer-employee relationship. Policy makers and researchers will also be critical in arguing for the creation of the data infrastructure, shaping its development to answer critical policy questions and fill evidence-gaps that government urgently needs to address.

We anticipate that all parties will seize the opportunity that we believe now presents itself for the UK to build a new LEED infrastructure for the public good.

Motivation for a new LEED infrastructure

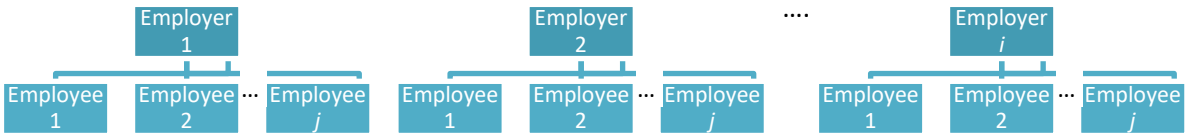
1.1 Introduction

Linked employer-employee data (LEED) are a valuable component of any country's data infrastructure, providing information on the interaction between workers and employers in the labour market across time and space. Their unique features enable LEED to provide a more comprehensive understanding of the dynamics of the economy and labour market – and the links between business performance and household living standards – than if one were to study organisations and households in isolation.

At the centre of any LEED are observations on employees linked to their employing unit (see Box 1). When these data offer comprehensive coverage of the economy and contain unique and time-consistent identity linkages between employees and employers, they can form a *linked employer-employee data spine,* which provides the basis for an extensive and highly-informative LEED data infrastructure.

Box 1: The basic features of linked employer-employee data

Linked employer-employee data (LEED) comprise data on employers and at least one or more of their employees. The employer unit may be a workplace ('local unit') or an organisation ('enterprise'). For enterprises occupying a single site, the workplace and the organisation are synonymous. The best available LEED offer data on employers and employees at multiple time points (longitudinal LEED).



The value of integrated LEED has long been recognised in many other OECD countries, where LEED forms a central part of the data infrastructure that governments and academics can leverage to understand the operation of labour markets and the economy. However, no such data infrastructure currently exists in the UK. This restricts our analytical capability and, ultimately, makes evidence-based policy making more difficult. It also means that the UK is sometimes absent from international comparisons.[1] We propose that the Economic Statistics Centre of Excellence (ESCoE) should collaborate with the Office for National Statistics (ONS) and HM Revenue and Customs (HMRC) over the next 12 months to develop a **prototype for a new LEED infrastructure**, consisting of a LEED spine linked to a small number of additional datasets.

[1] In the workshop on 'Developing New Linked Employer-Employee Data for the UK', held in London on 12th September 2024, Alexander Hijzen, Head of the OECD's LinkLEED network, stated that the UK is not a core part of the network's current program of work because of the limitations of the UK's LEED infrastructure.

HMRC's sharing of employment and earnings data from its Pay-As-You-Earn Real Time Information (PAYE RTI) system with the ONS provides a unique and exciting opportunity to build a new LEED infrastructure for the UK. Specifically, the PAYE RTI data allow for the creation of the UK's first ***linked employer-employee data spine***: a canonical, longitudinal reference database linking all payrolled employees to the organisational units in which they are employed across space and time. The spine itself would have useful analytical potential. It would also constitute an ***enabling infrastructure***, providing a consistent frame through which a variety of existing employee and firm-level administrative and survey datasets could be linked to one another, and acting as a sampling frame for new employer and employee-level surveys.

LEED of this nature would have many research uses with high potential to impact policy making. They would be valuable to government stakeholders and academic researchers with interests in the functioning of the economy and labour market. They would enable the publication of new statistics on labour market dynamics, complementing existing measures and supporting the long-term development of labour market statistics in the UK. This could include using items from LEED to substitute for survey questions in the Labour Force Survey (LFS) or the Annual Survey of Hours and Earnings (ASHE). They would also allow stakeholders to fill evidence gaps on important questions for policy makers in areas such as economic growth, clean energy and barriers to opportunity (see Table 1). Specific illustrations are provided in Sections 2-4.

Table 1: LEED and the government's missions

Economic Growth
• Who benefits from economic growth? • How do skills and other employee attributes contribute to firm productivity? • How do changes in labour market regulation affect business performance?
Clean Energy
• What are the impacts of decarbonisation on employees? • What role do employees play in diffusing green innovation between firms?
Economic Opportunity
• What are the lifecycle impacts of working at a high or low productivity firm on earnings? • What are the risk factors associated with job loss and inactivity?

To build the prototype of the LEED spine, mentioned above, we propose that ESCoE, ONS and HMRC should utilise the PAYE RTI data provided by HMRC to ONS under a Memorandum of Understanding (MoU).[2] In this prototype, linkages should be established to a small number of additional datasets held by ONS to demonstrate the spine's potential, including (but not restricted to) the Longitudinal Business Database. This prototype infrastructure should be used to generate short-term outputs highlighting the value of LEED. Collaborations should be sought with other stakeholders, such as Administrative Data Research UK (ADR UK) and the Economic and Social Research Council (ESRC), in order to develop the prototype into a stable, regularly-updated and integrated LEED infrastructure, with the potential for additional data linkages and with the potential for the spine to be used as a sampling frame for new surveys. In parallel, we propose that ONS

[2] This MoU operates under the permissive gateway created by the Statistics Registration Services Act (as amended by the Digital Economy Act 2017).

and HMRC should seek to ensure that the LEED spine and surrounding infrastructure can be made available within a trusted research environment for use by the wider research community.

Some moves towards wider data sharing have already been made with the depositing by ONS and HMRC of a monthly PAYE RTI dataset with the Integrated Data Service (IDS) (see Integrated Data Service, 2025). This dataset has been deposited in the IDS as a means of making available the data underlying the experimental employment and earnings statistics published by ONS in collaboration with HMRC (see Leaker and Taylor, 2019). It covers the period January 2017 to December 2023 and is indexed to the Demographic and Business Indices that form part of the ONS Reference Data Management Framework (RDMF). There is an expectation that the dataset will be revised as data updates become available.

This dataset could be one starting point for the development of a new LEED infrastructure, using the promised capabilities of the IDS to effect links to other datasets held on the platform. However, the standardisation around monthly observations will not suit all needs, the linkage capabilities offered by RDMF indexation are still largely unproven and the IDS is still in a developmental state. Any IDS-based data asset is also restricted to analytical uses; the IDS dataset does not contain the identifying information that would be needed for it to function as a sampling frame. So our view is that the PAYE RTI IDS dataset is best seen as one potential realisation of the LEED spine. We propose that ONS and HMRC should incorporate the efforts involved in creating and depositing this dataset within a broader initiative to establish a ***comprehensive and flexible LEED infrastructure*** which will meet a variety of needs. We expand on this proposal in Section 1.4. First, however, we discuss the value of LEED for evidence-based policy making and the UK's international position.

1.2 The value of LEED for evidence-based policy making

LEED are uniquely valuable for a number of reasons. First, they provide observations from both the supply and demand side of the employment relationship. This means that LEED can provide data on employee jobs and remuneration which is consistent with information on employers. This is a key advantage in the production of National Accounts (United Nations, 2025). The combination of data from employers and their employees also makes it possible to answer research questions that are not easily addressed with employee-only or employer-only datasets, where the set of observed characteristics is typically one-sided. For example, a rich analysis of the drivers of firm productivity requires information on management practices, and on employees' skills and attitudes to work, to see inside the "black box" linking employer and employee inputs to firm performance (see Bender et al, 2018).

Longitudinal LEED also allow for the publication of new economic statistics on labour market dynamics, including hiring rates, separation rates and job-to-job flows (e.g. Abowd et al, 2005). They also facilitate analysis of the interactions between labour market dynamics and business dynamics (e.g. Davis et al, 2006; Upward and Wright, 2019), the factors which support job stability and career progression (e.g. Fredericksen et al, 2016) and the importance of worker mobility for outcomes at both the worker and firm level (e.g. Casarico and Lattanzio, 2024). One can also investigate the antecedents and consequences of the adoption of particular management practices or technologies (such as AI). And one can investigate the implementation of regulations or the provision of business or worker support, permitting a more rigorous assessment of policy interventions than might otherwise be the case.

LEED also make it possible to understand the sources of heterogeneity in employees' experience of work in more detail. For instance, one can determine the relative importance of firm and employee attributes in shaping wage inequality or job insecurity (e.g. Criscuolo et al, 2023a; Henning and Stadler, 2023). This

perspective is often missing from employee-only datasets, which may then attribute firm-level variance to a person's demographic traits, occupation or region (see Phan et al, 2023).

1.3 The UK's position internationally

Across the OECD, policy makers and researchers rely on LEED to provide insights into the operation of labour markets and the employment relationship. It is common for the National Statistical Institute to re-purpose administrative data which is available as a by-product of government activity such as taxation, social security transfers or other interactions between government and organisations or citizens. Many countries also supplement these administrative data with large-scale survey data drawn from employers and/or employees.

Three countries – the United States, Denmark and Germany – provide an indication of what is feasible.

In the United States, the *Longitudinal Employer-Household Dynamics (LEHD) dataset* comprises a quarterly database of linked employer-employee data covering all wage earners. The core of the LEHD consists of data on jobs and earnings. Employee demographic characteristics and employer characteristics are matched in from other sources, along with annual information on firm sales and productivity, and data on firms' management practices

The Danish *Integrated Database for Labor Market Research (IDA)* is an annual database combining information on the universe of labour force participants and workplaces in Denmark. The IDA is compiled from a variety of registers maintained by Statistics Denmark. Data on annual hourly earnings is augmented with employee characteristics from the Population Register and Education Register. Employer characteristics and annual accounts data are linked in from various firm-level registers (e.g. FIRE, FIRM).

The German *Linked Employer-Employee Data (LIAB)* combines individual employment biographies generated from the social security system with establishment data collected from administrative sources and an annual establishment panel survey. Since 2012, the IAB has also used the LIAB as the basis for additional, periodic linked employer-employee surveys which collect additional data on human resource practices, employee work attitudes and wellbeing.

Further information on the LEHD, IDA and LIAB is provided in Appendix A. Summary information on the LEED that is available in other countries – including France, Italy, Portugal, Norway, Sweden and New Zealand among others – is provided in the appendices to various OECD publications (e.g. Criscuolo et al, 2021, 2022, 2023a, 2023b).

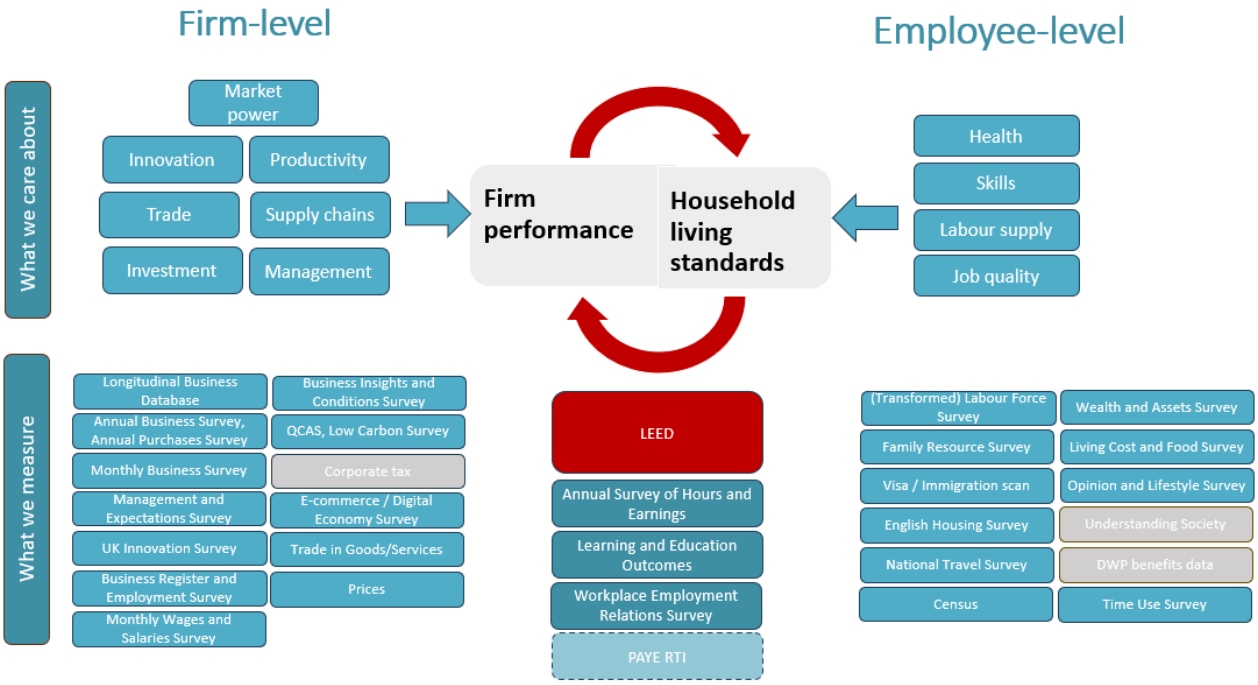
These various LEED infrastructures have a number of common features:

- The data infrastructure is **based on a linked employer-employee spine** that gives comprehensive coverage of employees and their employing organisational units
- This spine contains **unique and time-consistent identity linkages** for employees and employers to allow workers and firms to be tracked over time
- The spine is **augmented** with additional employee and employer characteristics, matched in from administrative databases and surveys
- The data are **periodically updated and curated** in a transparent and replicable way by the National Statistical Institute (or government agency in the case of the IAB)
- The data are made available to the research community as a **public good** under controlled conditions.

Such an infrastructure does not currently exist in the UK. There are many sources of data which provide insights into the supply and demand sides of the labour market (Figure 1), but few of these are linked employer-employee datasets.

In terms of LEED, evidence based policy-making and evaluation in the UK is currently heavily reliant upon the *Annual Survey of Hours and Earnings (ASHE)*. The ASHE is based on a 1% sample of employee jobs, with employees selected into the sample via the last two digits of their National Insurance number (NINo). The data are longitudinal, since employees appear in the issued sample in each year that they hold an employee job, with a unique personal identifier allowing observations for the same employee to be linked over time. The data also contain an Inter-Departmental Business Register (IDBR) enterprise reference number, enabling analysts to identify employee observations from the same firm. The ASHE data have been augmented via linkages to other datasets, notably by the Wage and Employment Dynamics (WED) project team. However, with only a 1% sample of employees, ASHE is very limited in what it can say about a single firm's workforce. The reliance on annual observations and the relatively high rates of longitudinal attrition (Forth et al, 2024) also limit what the data can say about employment and earnings dynamics at the individual level.

Figure 1: Key sources of UK data on firms and employees



The DfE's *Learning and Education Outcomes dataset (LEO)* is designed to provide insights into the long-term effects of education on people's lives, particularly their employment, earnings, and benefits. It aims to help researchers and policymakers understand the pathways people take through education and into the labour market. LEO links together de-identified data on individuals from various sources including: education records (from schools, colleges, and universities); employment and earnings information (from HMRC); and benefits information (from DWP). Employer information is matched in at enterprise-level from the IDBR using PAYE

employer references. These employer references allow one to identify employees in the same firm. However, LEO does not give comprehensive coverage of the labour market, including only those people who were either: (a) born since 1985 and have engaged with the state school education system in England (plus those attending independent schools, who are included in the Key Stage 4 and Key Stage 5 data); or (b) were born before 1985 and have attended an English further education institution since 2002/03. The employer information is also limited to characteristics such as size, industry and legal status.

A third source of LEED comes from the series of *Workplace Employment Relations Surveys (WERS)*, although this series has not been updated for over a decade. WERS was initially a survey of workplaces only, conducted in 1980, 1984 and 1990, but in its last three iterations (1998, 2004 and 2011) it supplemented the survey of workplace managers with a random sub-sample of up to 25 employees from those workplaces. Historically, the WERS data provided a critical source of evidence for policy makers when examining the implications of new legislation in areas such as the minimum wage, trade union recognition, employee information and consultation, equal opportunities and flexible working (Drew et al, 2013). However, the use of the workplace as the unit of observation made it difficult to collect or match on accounting-type data on business performance, thus limiting the survey's contribution to the productivity debate. The survey was also challenging to administer, because of the reliance on the employer to facilitate the sampling of employees. The sample design, and the survey's use of face-to-face interviewing for managers also made it resource-intensive, meaning that it was repeated on a long cycle, which limited opportunities for longitudinal analysis.

The shortcomings of existing sources of LEED mean that the UK currently has a significant gap in its statistical and research infrastructure. This means that economic and social policy making is either based on UK LEED with significant limitations, or it is based on insights developed from LEED in other settings, namely the US and Europe. However, the UK now has a strategic opportunity to get back to the frontier by developing new LEED based around the PAYE RTI data: the fourth source of LEED indicated in Figure 1.

1.4 A strategic opportunity

Historically, the PAYE RTI data have only been available for research within the HMRC Data Lab, where they can only be linked to other HMRC datasets. However, the data are now becoming more widely available under the provisions of the Digital Economy Act (2017). In particular, PAYE RTI data are now routinely shared with ONS under the MoU discussed in Section 1.1. This allows ONS to use the data to generate experimental earnings and employment statistics (Leaker and Taylor, 2019) and also for wider statistical and research purposes (e.g. ONS, 2024), since the MoU allows ONS to link PAYE RTI to other specified ONS datasets. As noted above, one version of the data has recently been deposited with the IDS (Integrated Data Service, 2025).[3]

The PAYE RTI overcome several of the limitations of existing sources of LEED in the UK:

- They are **comprehensive in coverage**: an employer must register for PAYE if they pay at least one employee at or above the lower earnings limit for National Insurance contributions (£123 per week for the 2024/25 tax year). Registered employers must then supply HMRC with information on each employee on their payroll (including those below the lower earnings limit). This generates a set of

[3] Extracts from PAYE RTI are also provided by HMRC to DfE to augment the LEO dataset. Similarly, employee-level extracts are provided to DWP for the construction of their Employment Characteristics and RAPID databases.

records covering almost all employee jobs in the UK, linked to the PAYE schemes under which they are registered.

- They are **longitudinal**: Employees are uniquely identified by their NINo. Employers are uniquely identified by the PAYE scheme reference number, which can be linked within ONS to the unique enterprise reference numbers (Entrefs) used on the Inter-Departmental Business Register.[4] These identifiers enable employee records to be grouped within employing units and linked across time.
- They are **high-frequency** and **regularly updated**: Employers are required to submit information to HMRC each time they pay their employees. The PAYE RTI dataset collates these 'payslips', most of which record either weekly or monthly information on an employee's gross earnings, income tax deductions, National Insurance contributions and pension contributions. ONS holds data from April 2014 onwards and continues to receive regular updates from HMRC.
- They are **linkable** to other datasets: Whilst the data provided within the PAYE RTI dataset itself is relatively limited, the availability of unique employer identifiers (PAYE references) and employee identifiers (NINos) enables the PAYE RTI data to be used to create a "LEED spine": a series of reusable longitudinal identity linkages. The key identification variables then enable linkages to be made with other datasets, either directly , or indirectly via the ONS demographic and business indices that form part of the RDMF.[5]

Use of the data is necessarily restricted. The current MoU between ONS and HMRC allows HMRC to share identifiable PAYE RTI data with ONS for the purposes of supporting ONS' functions. These functions may include the production of statistics or research within ONS. The MoU does not permit data sharing outside ONS. However, Part 5, Chapter 5, Section 64 of the Digital Economy Act (2017) provides a permissive legal gateway to facilitate the sharing of de-identified data by ONS and HMRC for accredited research purposes, on the condition that any such data is made available only to accredited researchers in a secure environment. It is these DEA research powers that have been used to facilitate the recent deposit of PAYE RTI data in the IDS.

1.5 Our proposal

In our view, the roadmap for the development of a new LEED infrastructure comprises three related stages:

- **Stage 1:** Utilise the PAYE RTI data to develop a **LEED spine** which will provide unique and time-consistent identity linkages for employees and employers, allowing workers and firms to be tracked over time
- **Stage 2:** Develop **linkages** from the LEED spine to a range of existing administrative and survey data sources, so that additional data on employers and employees can be added to the spine
- **Stage 3:** Use the spine as a **sampling frame** for new employer-employee surveys, so that gaps in our existing measurement of the employer-employee relationship can be addressed.

To move forward, we propose that ESCoE, ONS and HMRC should collaborate to develop a prototype for a new LEED infrastructure, consisting of a LEED spine linked to a small number of additional datasets already held within ONS. In doing so, ONS should build on its unique position as the UK's National Statistical Institute with unparalleled capabilities to process and link disclosive datasets for the creation of research-ready data. This work would also build on ONS' and HMRC's experience with the PAYE RTI data (Leaker and Taylor, 2019;

[4] Some organisations operate more than one PAYE scheme and so that mapping is not one-to-one. See Section for a further discussion.

[5] The same concept of a LEED spine was central to the construction of the UK Longitudinal Business Database (LBD) (Lemma et al., 2023).

Integrated Data Service, 2025), their existing MoU, and ONS' experience in building the UK Longitudinal Business Database as a reference spine of businesses (Lemma et al, 2023). ESCoE would contribute academic expertise in the creation and analysis of LEED. Additional collaborative partnerships should also be sought with those outside ONS who have experience of using PAYE RTI data, such as the Wage and Employment Dynamics Project team, who have linked PAYE RTI to ASHE for research use (Ritchie et al, 2025), the Centre for the Analysis of Taxation, who are linking PAYE RTI to HMRC business datasets in the HMRC Datalab, and others who have used PAYE RTI to generate experimental labour market statistics (Dorsett and Hug, 2022).

If the prototype serves to solidify the case for investing further in new LEED for the UK, collaborations would then be required from other stakeholders, such as ADR UK and ESRC, in order to develop the prototype into a stable, regularly-updated and integrated LEED infrastructure, incorporating additional data linkages and with the potential for the spine to be used as a sampling frame for new surveys. The intention would be to use the DEA research powers (subject to HMRC consent) to make non-disclosive elements of this LEED infrastructure available for use by accredited researchers within trusted research environments. We propose that ONS would act as the custodian of this LEED infrastructure, maintaining and updating the infrastructure as new data become available. The over-arching intention would be to create a transparent, robust, replicable and valuable public good which can support research and policy making into the future: a major addition to the UK's wider data infrastructure.

1.6 Outline of the remainder of the report

The remainder of the report outlines the key features of a future LEED infrastructure for the UK in more detail and sets out a road map for the development of this infrastructure.

- Section 2 discusses the key features and of the LEED spine and outlines important considerations for its construction and use.
- Section 3 outlines the range of linkages that could be made between the LEED spine and other existing sources of administrative and survey data.
- Section 4 then discusses opportunities to use the LEED spine as the sampling frame for new surveys, in the spirit of WERS.
- Section 5 concludes by summarising the key features and capabilities of a new LEED infrastructure built around the PAYE RTI data, and by proposing a series of next steps.

Developing the LEED spine

2.1 Introduction

This section discusses the key features of the LEED spine, and outlines some of the main considerations in its construction.

At its core, a LEED spine would comprise a set of comprehensive, longitudinal identity linkages, connecting individual employees to their employing enterprises across successive time periods. These linkages may be thought of as identifying 'jobs', where a job is defined as a match between an individual employee and an individual employer at a specific time point.

Construction of the spine relies on PAYE RTI data as this is the only data source for the UK which can provide a comprehensive mapping of employee jobs to employers. It does so by providing regularly-updated information from employers on the payments they are making to employees within their PAYE scheme (or schemes). Employees are uniquely identified by NINos. The canonical LEED spine would be created by combining the NINo-PAYE linkages in the PAYE RTI dataset with the PAYE-Entref linkages contained within L-WEP, which is an ONS data table linking Enterprise Groups, Enterprises and PAYE schemes, and used in the construction of the LBD (Lemma et al., 2023).

The LEED spine would constitute a dynamic reference framework linking employees to their employers. However, it would also have important analytical capabilities. For instance, it could be used as a means of measuring labour dynamics, including rates of employee mobility across firms, and firm-level hiring and separation rates. The spine could also serve as flexible, reusable and consistent framework to link a wide range of employee-level and firm-level data from administrative and survey sources. And it could be used as the basis for employer-employee sampling. Table 2 summarises the key properties of a flexible and re-usable LEED spine, distinguishing it from an ad hoc linked employer-employee dataset compiled for a specific research project.

Table 2: Features of a data spine vs a linked dataset

Data spine	Ad hoc linked dataset
Primarily consists of identification information, including reference numbers, along with some basic characteristics of the job	Contains identifiers but primarily consists of variables that give additional information about the employee, employer or job, deriving from data linked to the spine
Repeatedly updated to capture longitudinality and continuity across the full series	One-off construction, making use of longitudinal or cross-sectional information, depending on the research question
Constructed transparently to be replicable and re-useable as the basis for different linkage projects	Tailored for one type of use, according to the research question in mind
Can be used as a sampling frame for future surveys	Constructed for secondary analysis only

Adapted from Lemma et al. (2023)

The remainder of the section first outlines the key features of the PAYE RTI data which make it valuable as the basis for a LEED spine; some limitations of the data are also identified. The section then goes on to discuss briefly how the PAYE RTI data would be processed to create the LEED spine. Finally, the section outlines some potential analytical uses. Section then goes on to discuss use of the spine as a framework for data linking, and Section discusses use of the spine as a sampling frame.

The PAYE RTI IDS dataset (Integrated Data Service, 2025) and other realisations of the PAYE RTI data, such as those described by Dorsett and Hug (2022) and Ritchie et al. (2024), provide useful reference points for the data processing tasks involved in creating a canonical LEED spine.

2.2 Input data and key features

An overview of the PAYE RTI data was provided in Section 1.4. The core content of the dataset consists of a series of records of payments paid by employers to their employees within their PAYE scheme (or schemes) (see Figure 2). The data is intended to capture all PAYE schemes where at least one employee earns above the Lower Earnings Limit for National Insurance and, in these schemes, all employees are covered. These core details are provided in a suite of three data files:

- *PAYE scheme file:* information on the PAYE scheme through which the employer makes payments to the employee
- *Employment file:* information on the employment (or job) that the employee holds with the employer
- *Payments file:* information on payments made to the employee within the job

A number of ancillary files then provide additional details. For instance, one data file provides some basic demographic information about the employee, such as their gender and date of birth. Another data file provides an end-of-year summary of all payments made in a job over the previous tax year.

The Payments file provides information on:

- All pay, including wages, salaries, fees, overtime, bonuses and commission
- All statutory payments (e.g. Statutory Maternity Pay, Statutory Paternity Pay)
- Benefits in kind, expenses and redundancy payments that are paid via payroll
- Payments from registered occupational pension schemes.[6]

The payment data do not capture any benefits, expenses and payments that are not paid via the payroll. They also exclude stock options not paid through payroll, employer National Insurance contributions and employer contributions to pension schemes.

The nature of the pay period varies between employees. Most are paid either monthly or weekly, but the data also allow for fortnightly, quarterly and other intervals, as well as one-off payments (see Table 3). The mixture of payment periods requires some standardisation: this process – termed 'calendarisation' – is outlined briefly in Section 2.3.1.

Further information on the content of PAYE RTI is provided in HMRC's RTI Data Item Guide (HMRC, 2023), although the data files provided by HMRC to ONS and others typically contain a restricted set of variables.

[6] It is generally not possible to separate components of wages (e.g. to identify the value of bonuses); however, benefits in kind and pension income can be distinguished from earnings.

Figure 2: Basic structure of PAYE RTI data

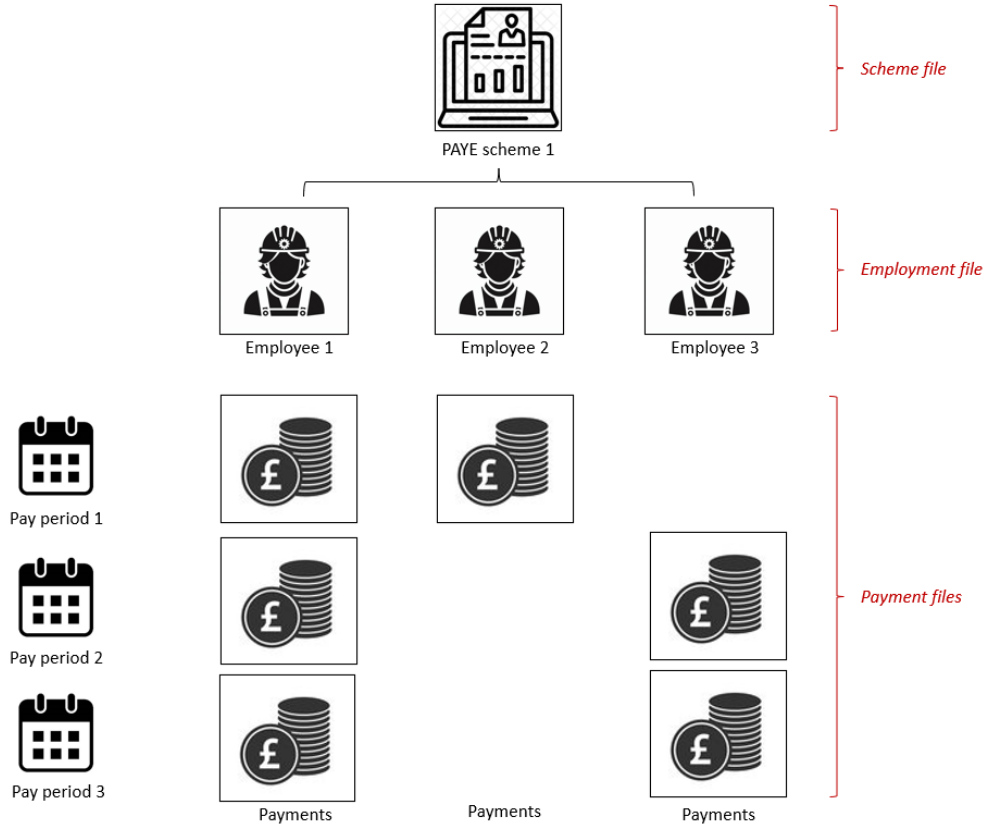


Table 3: Payment periods in PAYE RTI

Code	Payment period
W1	Weekly
W2	Fortnightly
W4	4 Weekly
M1	Calendar Monthly
M3	Quarterly
M6	Bi-annually
MA	Annually
IO	One-off[7]
IR	Irregular

[7] A 'one-off' payment is where someone is employed to do a one-off piece of work, say, for one week or month, and only receives one payment. This differs from the other pay frequencies where an employee remains in your employment although paid on an irregular, quarterly or annual basis.

The main advantages of these PAYE RTI data over other sources such as ASHE or the LFS were alluded to in Section . Specifically, the data are:

- **Comprehensive in coverage**: covering all payrolled employees and their jobs in the UK
- **Longitudinal**: such that employee records can be linked across time from a starting point in April 2014
- **High-frequency** and **regularly updated**: with employers required to submit information to HMRC each week or month they pay their employees, and with ONS continuing to receive regular updates from HMRC.
- **Linkable:** such that one can identify all employees paid within a specific PAYE scheme, and such that employees and PAYE schemes can be linked to other employee and employer datasets.

The data do have a number of limitations, however:

- **Limited data on hours worked:** Employers are asked to supply information on the employee's normal weekly working hours on each employment record, coded to one of five categories (Up to 15.99 hours; 16-23.99 hours; 24-29.99 hours; 30 hours or more; Other – if the employee does not have a regular working pattern or if the payment represents pension income). This categorical information was originally designed to assist with eligibility for Working Tax Credits. However, it prohibits the derivation of hourly earnings (e.g. for the purposes of identifying proximity to the National Living Wage) and the identification of zero-hours contracts.[8]
- **No data on occupation:** The data contain no occupational classification. This must be linked in from other sources, where available (see Section 3).
- **No establishment identifier:** In organisations with multiple establishments (workplaces), there is no information which connects employees to a specific establishment. Workplace location must therefore be derived (see Section 2.3.1).

These limitations are not unique to PAYE RTI. For instance, the US LEHD lacks data on hours, occupation and workplace location.

2.3 Creating the LEED spine

2.3.1 Processing the PAYE RTI data

The primary objective in processing the PAYE RTI data is to create a reference dataset which connects individual employees to PAYE schemes across successive time periods. Since individual employees are identified via unique National Insurance numbers (NINos) and PAYE schemes are identified via unique PAYE reference numbers (PAYErefs), this involves the creation of a longitudinal NINo*PAYEref panel.

The obvious starting point for creating the NINo*PAYEref panel is the Employment data file, which notionally holds start and end dates for each job. However, in practice, observed earnings in the Payments data file are sometimes inconsistent with these dates, with payments sometimes being observed before the recorded start of the job or after the recorded end. The derivation of job spells therefore requires some imputation using data from both the Employment file and the Payments file. With this in hand, however, one is able to identify job matches and their duration, which in turn allows for the creation of a NINo*PAYEref panel at any chosen

[8] A proposal to require employers to report the number of paid hours worked by an employee within each pay period was dropped in January 2025 (HMRC, 2025).

frequency (daily, weekly, monthly, quarterly, annual etc). For example, a quarterly panel would align with the periodicity in the LBD.

As shown in Table 3, payments in PAYE RTI may refer to a variety of different pay periods. Around three-fifths of jobs are paid monthly and around one quarter are paid weekly (Integrated Data Service, 2025), with the remainder paid at other frequencies (e.g. fortnightly). In order to create a panel with a common unit of time (e.g. a quarterly panel) it is necessary to put the payments recorded in PAYE RTI onto a consistent time-basis via a process of calendarisation. The payments recorded in PAYE RTI are first converted into daily employment pay rates for the estimated period of employment over which the work was done. These daily pay rates are then multiplied by the estimated number of days worked in the period on which the panel is based (in this example, quarter-by-quarter).[9]

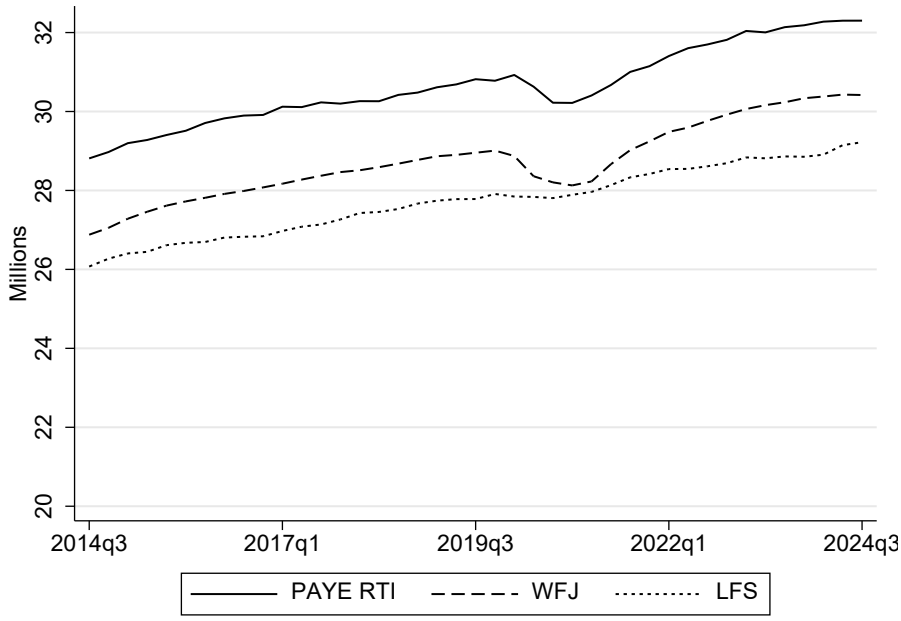
These approaches, and other nuances of processing PAYE RTI, are set out in more detail in ONS (2022a), Dorsett and Hug (2022: Appendix A) and Ritchie et al (2024).[10]

There are some differences between PAYE RTI and other established sources of statistics in terms of coverage and content (see Leaker, 2022). For instance, PAYE RTI includes anyone who has a paid employee job, whereas the ONS Workforce Jobs (WFJ) series and LFS statistics include only those aged 16 or over. PAYE RTI includes pension income, which would not be included in measures of earnings captured in other ONS series. However, the RTI series tracks the employee jobs (EJ) element of WFJ closely (see Figure 3). And after removing those records which are judged to refer to occupational pension payments, the level and growth rate of weekly pay in the PAYE RTI data is similar to that shown in the Average Weekly Earnings (AWE) series (Figure 4) and the distribution of weekly pay is very similar to the distribution found in ASHE (see Figure 5).

[9] This is preferable to simply summing payment amounts over a particular period, which can be problematic when aggregating data for jobs paid monthly and jobs paid weekly, as the number of weeks can vary in a month.

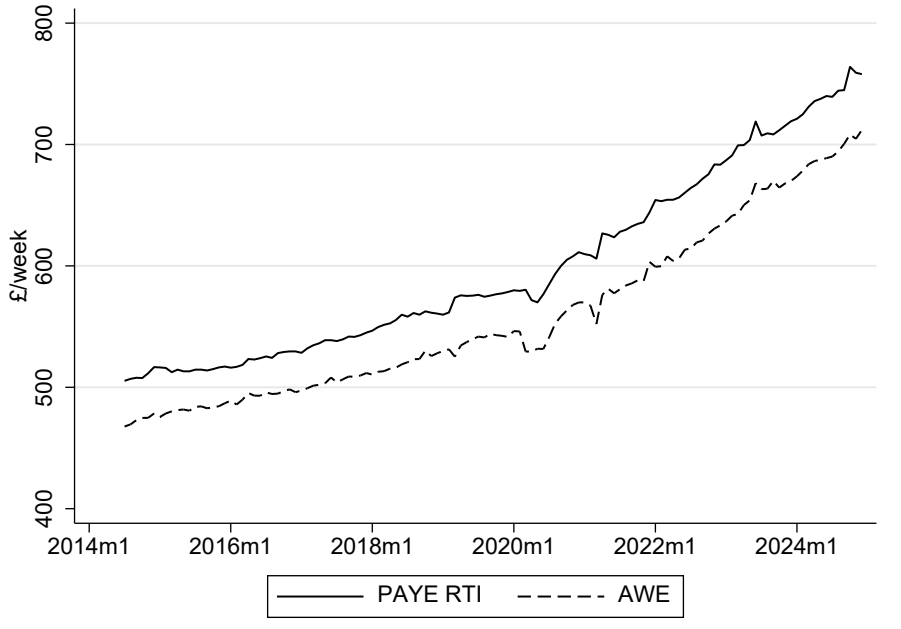
[10] The PAYE RTI IDS User Guide contains little detail on how the data have been processed for deposit in the IDS, but one can infer some details from ONS methodological publications relating to their experimental employment and earnings statistics, such as ONS (2022a).

Figure 3: Number of employee jobs from PAYE RTI, Workforce Jobs (WFJ) and the Labour Force Survey (LFS)



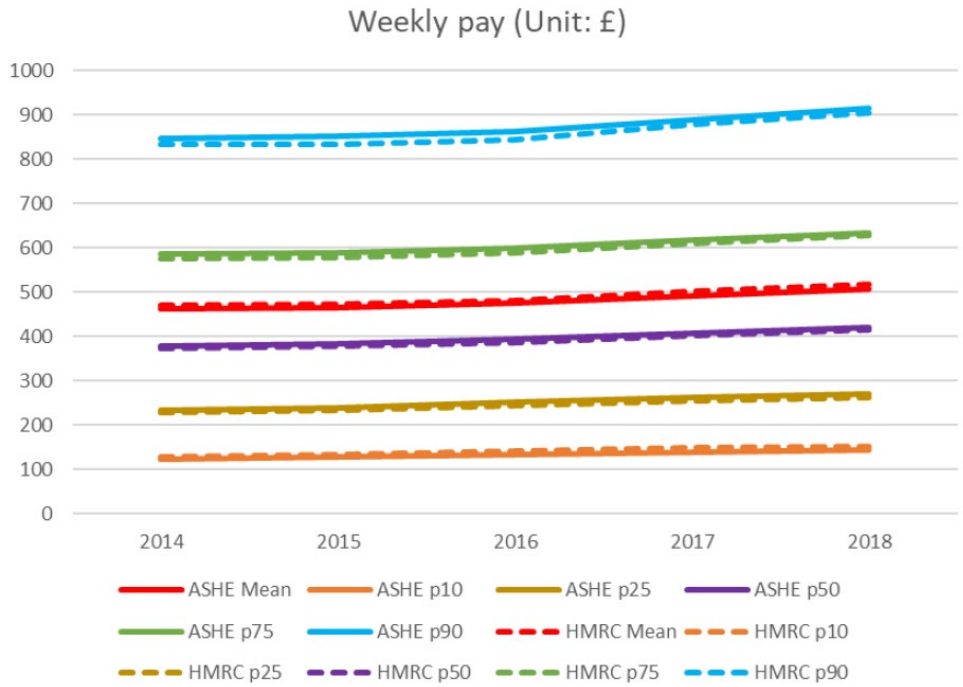
Notes: q1 = January-March; q3 = July-September. Values are seasonally adjusted. WFJ counts the number of employees, aged 16 years and over, that an organisation directly pays from its payroll; collected through a survey with a specific count date (source ONS series BCAJ). RTI counts anyone who has a paid employee job during the reference period; it is a monthly average of daily counts of the number of payrolled employees (source: ONS, 2025a). LFS counts the number of people aged 16 years and over in employee jobs (not the total number of jobs held) (source: ONS series MGRN).

Figure 4: Average (mean) weekly pay from PAYE RTI and Average Weekly Earnings (AWE)



Notes: Seasonally adjusted. The PAYE RTI series (source: ONS, 2025a) covers Northern Ireland, HM Armed Forces and government-supported trainees paid via PAYE, pay-rolled redundancy payments and signing-on fees, all of which are excluded from AWE (source: ONS series KAB9). We convert the published monthly PAYE RTI figure to weekly for comparison purposes.

Figure 5: Distribution of weekly pay in PAYE RTI and ASHE



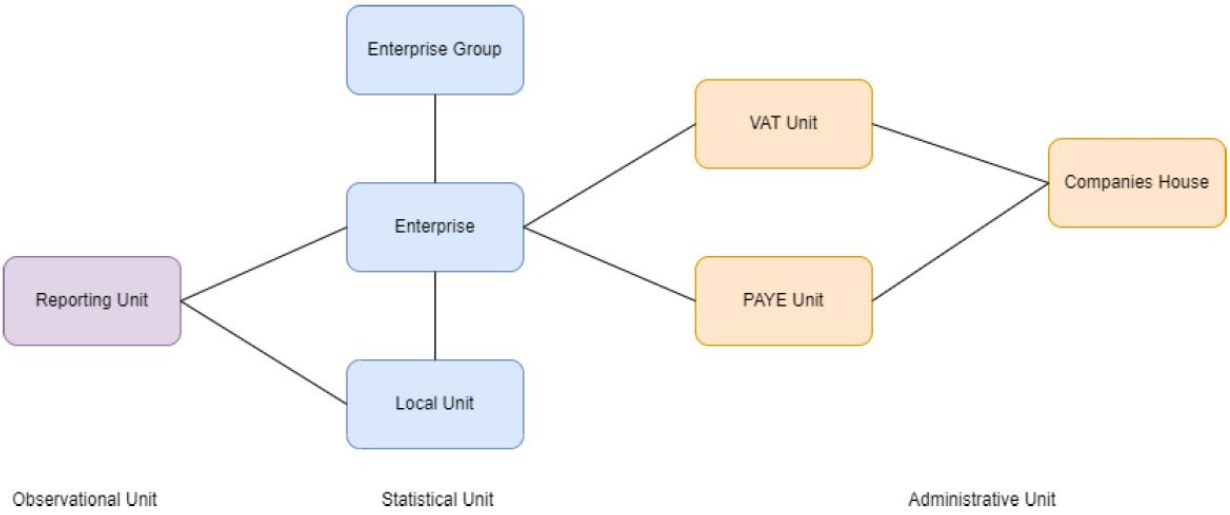
Notes: Compares weekly pay in HMRC PAYE RTI data and ASHE at the point in time of the ASHE survey (April) for employees with only one job. p10=10th percentile and so on. Source: Ritchie et al. (2025)

2.3.2 The employer dimension of the LEED spine

The process outlined above generates a NINo*PAYEref panel. However, the PAYE scheme is an administrative unit, the boundaries of which are determined by the employer. Some employers run multiple schemes for different groups of staff (e.g. monthly paid and weekly paid staff). In some cases, an employer may pay the same employee via different PAYE schemes (e.g. one scheme for regular earnings, another for bonuses). These decisions are not standardised across employers. Moreover, the PAYE unit is not a recognised statistical unit, and not a unit commonly used in other business datasets. So it has significant limitations as an end point for reference purposes, and also as a basis for analytical work and data linking.

Figure provides a basic illustration of the different units that may be used to identify a business (or part of a business). For statistical purposes, it is optimal for the employer identifier to be based upon a recognised statistical unit. However, in multi-site businesses, the PAYE RTI data contain no data on the establishment (local unit) at which the employee works. The natural employer unit for the LEED spine is therefore the enterprise. An enterprise is defined as the smallest combination of legal units that is an organizational unit producing goods or services, with a certain degree of autonomy in decision-making, especially regarding resource allocation.

Figure 6: Business structure on the IDBR



Source: Lemma (2023)

Enterprise Reference Numbers (Entrefs) can be attached to PAYE schemes by using the L-WEP table held by ONS and used in the construction of the LBD (Lemma et al., 2023). The L-WEP table acts as a reference table for all longitudinal links between Enterprise Groups, Enterprises and PAYE schemes. Once the Entref has been added to the NINo*PAYEref panel, it will be possible to aggregate an employee's pay and employments across multiple PAYE schemes within the same Enterprise (if there are multiple schemes), to create a ***job-level panel (NINo*Entref panel)*** which would form the basis of the LEED spine. An important criterion here is that the longitudinal spine of Entrefs should be consistent with records of firm entry and exit in the LBD.[11]

Column 2 of Table 4 shows the distribution of employees according to the number of PAYE schemes in the enterprise. It implies that some aggregation across schemes may be needed for up to one-third of all employees in the PAYE RTI data, but this represents the extreme scenario where different schemes are used to issue different types of payment to the same employee; more likely, different schemes are used for different types of employees. For those employees paid via multiple PAYE schemes in the same enterprise, the summing of pay across schemes within a given NINo*Entref pairing is straightforward. However, some rules may need to be devised to address cases in which the multiple PAYE records in such a pairing offer conflicting information on the nature of employment (e.g. on the start date).

[11] One minor complication in creating the Entref-level panel is that, in some cases, multiple employers use a single PAYE scheme to report PAYE information to HMRC. However, is not accepting requests for new "pooled PAYE schemes" and the remaining schemes are only used by a small number of employers.

Table 4: Distribution of employees across enterprises with specific numbers of PAYE schemes or Local Units

Number of PAYE schemes or Local Units recorded on BSD	Distribution of employees according to number of PAYE schemes (column %)	Distribution of employees according to number of Local Units (column %)
0	0.1	17.0
1	67.7	22.9
2	14.8	6.5
3	5.6	3.3
4	3.5	2.6
5 or more	8.2	47.6
Total	100	100

Source: Business Structure Database (counting only the numbers of live PAYE schemes and live Local Units)

Local units are not identified in the PAYE RTI data and so the data provide no information on the site at which the employee works in the case of multi-establishment organisations. However, the data do contain information on the employee's residential location. This information can be used to assign employees to local units probabilistically, by combining the residential location with information about the location of an enterprise's local units (taken from the IDBR) and data on travel times. The LEHD provides a template for such a worker-to-establishment imputation procedure (see Abowd et al, 2005: 174-179). This could be translated to PAYE RTI and validated using data from ASHE or the Census, both of which collect data on an employee's residential and workplace location. Imputed data on workplace location should then be sufficient to allow the LEED spine to generate granular place-based statistics.[12]

In the LEHD, the imputation procedure affects 30-40 per cent of employees (Abowd et al, 2005: 174). Our estimates from the BSD indicate that around 60 per cent of employees in the UK work in multi-establishment enterprises and could require imputation of their establishment address (see column 2 of Table 4). The figure may be higher in our case than in the LEHD because the LEHD input data is state-specific; the imputation is then only required for employees belonging to enterprises with multiple establishments in the same state.

One final issue which arises at this point is that many ONS business datasets to which the LEED spine might be linked use the Reporting Unit, rather than the Enterprise, as the unit of observation. Examples include the Annual Business Survey and Management and Expectations Survey. We return to this issue in Section as part of the discussion of data linking.

2.4 Illustrative uses

The core purpose of the LEED spine is to act as a canonical reference framework, serving as the basis for linkage projects and new linked employer-employee surveys. However, the basic spine will also have immediate analytical uses as a stand-alone dataset.

[12] See Lane and Stephens (2006: 270-272) for one discussion of the application of LEED to regional analysis. One would not seek to use the imputed local unit as a basis for survey sampling because there is likely to be some measurement error for individual employees.

First, it can serve as the basis for generating new statistical indicators measuring worker flows (specifically hiring rates, separation rates and job-to-job flows). Such indicators will help to extend our understanding of **business dynamism** in the UK. See Use Case 1 in Appendix B for further details.

Second, by measuring the earnings implications of different types of worker flow, it can serve as the basis for new statistical indicators to aid our understanding of the **dynamics of wage growth** in the UK. See Use Case 2 in Appendix B.

We expect these statistical indicators would be of value to the ONS, HM Treasury, the Bank of England and the Department for Business and Trade, among others.

Linking additional data to the LEED spine

3.1 Introduction

The LEED spine offers comprehensive longitudinal coverage of the population of employees, but the range of data items is relatively narrow and, hence, sufficient to answer only a small set of research questions. Nevertheless, the LEED spine can act as a flexible, reusable and consistent framework for linking other data sources. The LEED spine would then serve as an **enabling infrastructure** through which one can link various sources of employee and firm-level data to provide new insights into the supply and demand sides of the labour market.

Combining data in this way would generate a new, enhanced research dataset that can be used to address research questions that cannot adequately be addressed with a single source. It is also possible to use one source to assess the quality of the other (see Jarmin, 2019; Benzavel et al, 2020). For Jarmin (2019: 165), this integration of different sources of administrative and survey data is the hallmark of a 21st century statistical infrastructure.[13]

In this section, we discuss how the LEED spine could be augmented by linking to other existing sources of administrative and survey data. Such linkages rely on a number of enabling conditions:

- *Reliable linkage methodologies:* In some cases, it will be possible to link the employer and employee records in the LEED spine to other data sources by direct record linkage, using unique, common identifiers (NINos in the case of employees; Entrefs in the case of employers). Many datasets do not contain these identifiers. In the past, this would have made linkage difficult. However, the ONS Research Data Management Framework is providing new opportunities to link datasets that do not share common, unique identifiers (see Box 4).

- *Management of disclosure risk:* The sharing of data for research in the UK rests on the "five safes" principle (safe people, safe projects, safe settings, safe data and safe outputs). Linking datasets together necessarily raises the risk of disclosure by extending the set of data items that are observed for a given firm or worker, with the potential to make those data less 'safe'. However, in other countries with extensive LEED infrastructures, this is managed by providing researchers with specific cuts of the data, containing only the data items required for the analysis.

[13] Jarmin (2019) also includes unstructured digital data (from transactions or online interactions) within his framework. There has been progress in developing such sources for statistical use (e.g. Hansen et al, 2023), but generally their development is some way behind that of administrative and survey data sources, and so we do not consider them here for reasons of brevity. We recognise that they could contribute to a LEED infrastructure in the future.

- *Legal basis for data linking:* The existing MoU between the ONS and HMRC allows ONS to link the PAYE RTI data to a range of other datasets held within ONS, for the purposes of supporting ONS' functions. New consents would be needed from HMRC and other data owners if the PAYE RTI were to be linked to other datasets not covered by the MoU, and if any linked data assets were to be made available outside ONS for research purposes. It is not possible to pre-judge those decisions. However, the Digital Economy Act 2017 (DEA) provides a permissive legal gateway that facilitates the linking and sharing of de-identified data by public authorities for accredited research purposes (see ADR UK, 2023). The data owner must still give consent, but this is more likely to be given if the potential linkage is motivated by a strong use cases. We provide some illustrative use cases at the end of this section.

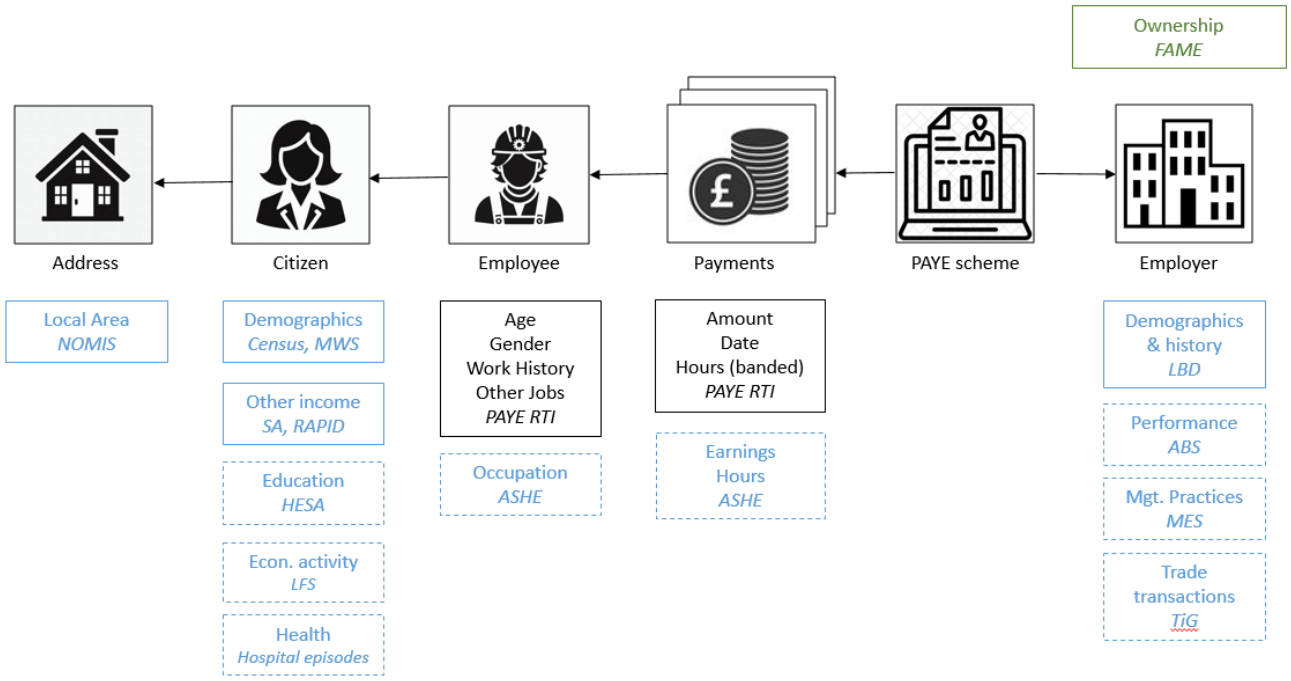
Box 2: The ONS Reference Data Management Framework

The Reference Data Management Framework (RDMF) is a collection of datasets that have been linked together to create a demographic index, location index and business index for the UK. These indexes contain one record for each person, location and organisation, respectively. Within each of these indices, each record has its own reference number (or index ID). In cases where one wishes to link two datasets which do not share a unique common identifier, the two datasets are first matched to the relevant index based on identifying variables. Those variables are then removed and the two datasets are linked via the index ID. ONS (2025d) provides further details.

Some of the potential linkages that could be made between the LEED spine and other datasets are illustrated in Figure . Section 3.2 then discusses some of the links to business data in more detail, whilst Section 3.3 discusses some of the links to employee data. Several illustrative use cases for linked LEED are outlined in Section 3.4.

Some linkages have already been made. The Wage and Employment Dynamics Project has linked a 1% subset of HMRC PAYE RTI records to ASHE and the HMRC Self Assessment (SA) data (Ritchie et al, 2024), with a further link to Migrant Worker Scan (MWS) data currently in progress. The Centre for the Analysis of Taxation (CenTax) has linked HMRC PAYE RTI data to a variety of HMRC business datasets to inform taxation policy for top earners. The ONS Health Research Group has linked PAYE RTI to the 2011 and 2021 Census datasets and to hospital episodes (ONS, 2024b, 2025c). ONS is also in the process of linking PAYE RTI to the Labour Force Survey (LFS). This implies that there is a great deal of experience to draw on. However, these linkage projects have been undertaken separately from one another. There is value in bringing this existing expertise together and making the LEED spine the focus of future linkage activity, to aid replicability and transparency and to avoid unnecessary duplication.

Figure 7: Map of potential linkages to the LEED spine



Notes on data sources: Black type = PAYE RTI; blue type = ONS/government datasets; green type = data outside government. Solid line = data likely to be available for *all* employees or employers; dashed line = likely to be available only for a *subset*. The abbreviated names of datasets (in italics) are expanded in the text.

3.2 Linking to business data

On the employer side, any linkages should seek to attach data on business demographics, performance and management practices. Potential datasets include:

Longitudinal Business Database (LBD) – Enterprise level: A quarterly business reference database built from consecutive snapshots of the IDBR which have been linked longitudinally to capture births, deaths and changes in activity. It provides quarterly data on such demographic events, along with enterprise turnover, employment, industry, location and legal status.

Annual Business Survey (ABS): An annual survey of businesses covering the production, construction, distribution and service industries. It provides indicators of economic activity such as the total value of sales and work completed by businesses, the value of purchases of goods, materials and services, stocks, capital expenditure, and total employment costs.

Management and Expectations Survey (MES): A periodic survey of production and service industries designed to gather information on the use of structured management practices and business expectations. Undertaken in 2017, 2020 and 2023.

Trade in Goods (TiG): Captures trade in good transactions reported to HMRC for administration and tax purposes. Contains information on the country of dispatch or destination, the value of trade and the commodity type.

Financial Analysis Made Easy (FAME): This commercial dataset contains information on companies registered at Companies House in the UK. It covers company financials (balance sheet, profit and loss account), along with information on the company's ownership structure.

One issue which arises in linking to many business datasets is that they are not collected at Enterprise level. Specifically, most ONS business datasets (including the ABS and MES) are collected at Reporting Unit (RU) level. FAME is constructed at the Company level. This does not preclude data linking, since ONS has reference tables linking Reporting Units and Company Registration Numbers to Enterprises (see Lemma, 2023). However, there is not a one-to-one correspondence between Enterprises and Reporting Units, or between Enterprises and Companies, in all cases.

Considering the case of an ONS business survey collected at RU level, such as the Annual Business Survey, there may be a need for some aggregation across RUs before data can be linked to the LEED spine at Enterprise level. Table 5 provides information on the number of RUs per Enterprise, weighted by employment. It suggests that this particular aggregation issue may affect around one third of employees, although this is an upper bound since it assumes that multiple RUs feature in the underlying dataset. If the dataset includes only one RU of a number within the Enterprise, one pragmatic approach is to take this RU to represent the whole.

Table 5: Distribution of employees according to the number of Reporting Units in the enterprise

Number of reporting units	Distribution of employees (column %)
0	0.1
1	63.7
2	19.1
3	6.7
4	4.0
5 or more	6.4
Total	100.0

Source: Business Structure Database

3.3 Linking to employee data

On the employer side, any linkages should seek to attach data on employee demographics, occupation, education and other sources of income. Potential datasets include:

2011 and 2021 Census of Population (Census): Provides data on (almost) the whole population of the UK on Census Day. Collects data on various demographic characteristics (ethnicity, disability, religion, country of birth) alongside educational attainment, job characteristics (occupation, hours), employer characteristics (industry, workplace location) and household circumstances.

Higher Education Statistics Agency (HESA) data on students: Covers students' engagement in courses that lead to a higher education qualification. Contains data on qualification type, level, subject(s) studied, whether completed and degree classification.

Migrant Worker Scan (MWS): A dataset produced by HMRC which identifies all NINos that are issued by application from foreign nationals. Any person who wants to take paid employment in the UK must have a valid NINO. Those who enter the country after the age of 16need to apply for a NINO and the data records their date of arrival in the UK, date of NINO issuance and nationality at time of application.

Self Assessment (SA) returns: A dataset produced by HMRC which contains data provided by individuals submitted a Self Assessment return for tax purposes. Provides data on the individual's income from employment, self-employment, investments, property and other sources relevant for the administration of the tax system.

Registration and Population Interaction Database (RAPID): Created by the DWP to provide a single coherent view of citizens' interactions with DWP and HMRC across the tax year. Provides information on benefit awards by type (including household-based benefits), tax credits and pensions in payment, alongside information on employment (from PAYE RTI), self-employment (from SA), migrant status (from MWS).

3.4 Illustrative uses

Appendix B provides a number of illustrative examples of how LEED could be linked to other data for policy analysis.

In Use Case 3, we outline the potential to use linked LEED for the evaluation of forthcoming changes to National Insurance Contributions (NICs) by linking the LEED spine to the LBD and ABS. We expect this use of LEED would primarily be of interest to HMRC, HM Treasury, the Low Pay Commission and the Bank of England.

In Use Case 4, we outline the potential to use linked LEED to analyse productivity and wage growth, also by linking the LEED spine to the LBD and ABS. We think this use of LEED would primarily be of interest to ONS, HM Treasury, the Department for Business and Trade, and the Bank of England.

In Use Case 5, we outline the potential for linked LEED to be used to identify the risk factors associated with economic inactivity by linking the LEED spine to the LBD, Census, HESA data, health data and RAPID. We expect this use of LEED would primarily be of interest to the Department for Work and Pensions.

In Use Case 6, we outline the potential for linked LEED to be used to investigate the "winners" and "losers" from trade, by linking the LEED spine to the LBD, Census, HESA data and trade data. We expect this use of LEED would primarily be of interest to the Department for Business and Trade.

3.5 A proposed roadmap for data linkage

As a starting point, we propose the ONS should seek to link the LEED spine to the LBD on the firm side and the 2011 and 2021 Censuses of population on the employee side. These datasets have the advantage that ONS is the data owner, and the data are also comprehensive in their coverage of firms and workers. This would add some basic demographic information about each enterprise and each employee to the LEED spine.[14] Such data are likely to be useful to all researchers using the LEED spine.

[14] The Census would only provide data for those in employee jobs in/around 2011 or 2021, but these data could be spread out over other periods in the panel.

We propose that a second stage of linking would focus on the Annual Business Survey (ABS), on the firm side, and qualifications data from the Higher Education Statistics Agency (HESA), on the employee side. Again, these data are likely to be of broad value.

Further linkages to other datasets could then be made at a later stage, depending on the particular research use that is prioritised.

Using the LEED spine as a sampling frame

4.1 Introduction

Thus far, we have discussed the linking of datasets that already exist, and which might be expected to continue to be available in the future. However, this suite of datasets is not completely sufficient for the analysis of labour market and employment relations issues in the UK. This is because many interesting aspects of the employment relationship are not recorded in administrative data or in existing surveys that are linkable to the aforementioned spine. Key examples include measures of corporate governance or leadership, aspects of firms' investments in intangibles, non-wage dimensions of job quality, and measures of employee engagement, commitment and well-being. Obtaining data on these issues involves running new surveys that can be linked into the wider data infrastructure by virtue of being sampled from the data spine.

In this section of the report, we discuss two specific ways in which the LEED spine might be used as a sampling frame for repeated employer-employee surveys that regularly add data to the existing infrastructure. One approach is to use the LEED spine as the basis for a new and improved version of the Workplace Employment Relations Survey (WERS), involving a new series of surveys on both the employer and employee side. The German Linked Personnel Panel (LPP) provides an obvious precedent (see Box 9). Another approach is to use the spine as the basis for adding an employee component to future waves of the ONS Management and Expectations Survey. These options cover similar ground, and might be seen as alternatives in some sense, although a WERS-type survey would likely cover a broader set of issues.

Box 3: The German Linked Personnel Panel (LPP)

The IAB in Germany has used their linked employer-employee spine to conduct periodic linked surveys of employers and employees – the Linked Personnel Panel (LPP) – which provides data resembling that previously collected in Britain via WERS. In the LPP, a subset of establishments participating in the IAB Establishment Panel have been selected and interviewed to collect more information on the establishment's HR practices. The establishments' employees have then been selected from the linked register and surveyed to collect information on their job characteristics, personality, work attitudes, organizational commitment and well-being. The LPP has thus far been conducted in 2012, 2014, 2016, 2019 and 2021 (Mackeben, 2023).

This is not the only potential use of a LEED spine for sampling, however. A LEED spine could also be used as the basis for targeted linked employer and employee surveys that address specific areas of government policy, such as changes to employment contracts, flexible working provisions or regulations governing information, consultation and trade union recognition. It could equally be used as the basis for linked surveys on developing issues, such as remote working and the use of Artificial Intelligence.

As in the case of linked data, the use of the LEED spine as a sampling frame comes with some specific considerations:

- *Legal basis for data sharing:* ONS would need to obtain HMRC's approval to use PAYE RTI data as a sampling frame, if this is not already covered by the existing MoU. Assuming that consent is provided, ONS would then be able to use employee details in the LEED spine as the basis for the employee

component of a linked employer-employee survey. Employer details are already under ONS control via the IDBR. If a suitable data sharing agreement were in place, ONS would also be able to pass sample information to a contractor to undertake any survey on ONS' behalf. Finally, HMRC would also need to provide consent to ONS for onward disclosure of the survey data for research purposes (under the DEA), since the employer-employee linkages derived from PAYE RTI would necessarily feature as part of the research dataset.

- *Burden on survey respondents:* The burden on survey respondents must be managed in any survey exercise. This burden may arise from the length of any survey or repeated survey requests. The former can be managed by ensuring that the survey does not attempt to re-capture information already available via data linking. The latter can be managed by centralised monitoring and control of survey requests, as is already done on the employer side via sample rotation for small businesses. Both are made easier by embedding new surveys within a broader LEED infrastructure.

4.2 Administering new WERS-type surveys of employers and employees

The WERS series has been commended as a "major and distinctive achievement of British social science" (Heery, 2014: 513). The series generated three linked employer-employee datasets in 1998, 2004 and 2011, and was valued by policy advisors, governments and academics alike, being used extensively in regulatory impact assessments, research and teaching.

The employer component of WERS was based on a sample of workplaces (local units) taken from the IDBR and comprised a wide-ranging interview with the senior manager responsible for employment relations issues, covering the governance and management of the workplace. For the employee component, the WERS interviewer obtained a list of all employees from the workplace manager and sampled 25 at random on site, issuing each employee with a self-completion questionnaire asking about the employees' skills, job tasks and attitudes to work.

The unusual approach to sampling employees is not thought replicable in the current environment. The employee data – being anonymised in the field – is also unlinkable to other data sources. However, a LEED spine would allow one to undertake regular new, inter-connected surveys of employers and employees – covering some of the topics that featured in the WERS surveys – and to do so in such a way that the resulting survey data was fully linkable to existing administrative (and survey) data on the sampled employers and their sampled employees.

In practical terms, sampling would be undertaken by first drawing a stratified random sample of enterprises from the IDBR to form the issued sample on the employer side.[15] Interviews would be conducted with the senior HR manager in the enterprise, conducted by telephone, video-call or push-to-web.

The LEED spine would then be used to sample a specified number of employees working at each of these sampled enterprises. The number of employees sampled in each enterprise could be fixed, so as to equalise the burden across workforces, or it could be set as some proportion of the workforce so as to capture a similar degree of variance within each workforce. The choice would ultimately be determined via piloting. Employees would be contacted directly at their home address (not via the enterprise) to avoid interference from

[15] This is an important, but necessary, break from the WERS series, which sampled at workplace (local unit) level. It is possible to sample at local unit level from the IDBR but, as discussed in Section 2.3.2, it is not possible to allocate employees to local units with certainly when creating the LEED spine.

managers (a potential weakness of the WERS approach), and could be invited to respond to the survey via post, telephone or push-to-web.

The advantage of using the LEED spine, apart from sampling, is that a range of data items will be available for many of the employers and employees in the spine via data linking. For instance, on the employer side, many businesses will have accounting-type data on firm performance available in the ABS. On the employee side, one will have earnings data (and work histories) from the PAYE RTI, and demographic information available from various sources. This will reduce the number of data items that need to be collected in the surveys.

The surveys would substantially enhance the range of data items available in the LEED infrastructure. The kinds of issues that could be covered are summarised below, based on the coverage of the WERS management and employee questionnaires. Any new questionnaires would be designed with a maximum completion time of 25-30 minutes.

The value of these data for investigating issues such as productivity and job quality are such that we would propose that such surveys are repeated every 3-5 years. Illustrative use cases are provided at the end of the section.

Table 6: Potential coverage of new WERS-type surveys

Employer survey	Employee survey
• Ownership characteristics and networks • Recruitment practices • Training and skill development practices • Team working • Task allocation • Flexible working arrangements • Performance management practices • Information and consultation practices • Employee representation • Payment systems and pay determination • Dispute resolution procedures • Equality, diversity and inclusion practices • Use of technology and AI • Values and corporate culture	• Contractual status • Structure of earnings • Job demands and working hours • Flexibility and working time quality • Autonomy and employee involvement • Promotion prospects and job security • Skill development and utilisation • Trust and fair treatment • Job satisfaction and well-being • Employee engagement • Organisational commitment • Technology use including AI

4.3 Adding an employee component to the Management and Expectations Survey

There is some overlap between the issues that might be covered in a WERS-type employer survey and the issues covered by the existing ONS Management and Expectations Survey (MES). So one might alternatively consider whether the LEED spine could be used to add an employee component to the MES.

The MES samples firms with 10 or more employees *excluding* firms in agriculture, financial services, and the public sector. In 2023, the issued sample comprised around 53,000 firms with an achieved response rate of 26.9%, implying around 14,000 responses. In 2024 ONS piloted a Public Sector Management Practices Survey, covering similar issues, with a response rate of 18%.

It should be recognised that the topics covered in the MES are somewhat narrower than those covered in WERS. Table 7 provides an overview of the coverage of MES, for comparison with the first column of Table 6. Crucially, WERS covers aspects of team production, employee involvement and workplace organisation which do not appear in the *MES*, but which are seen as important drivers of performance in other parts of the literature (Lawler, 1986; Huselid et al., 1995; Powell, 1995). However, if the MES is to continue to form part of the UK's survey infrastructure, adding an employee component to this repeating employer survey is perhaps a more pragmatic approach to generating rich survey-based LEED than instituting an entirely new linked survey series. Some broadening of the MES questionnaire would help to square the circle.

Table 7: Coverage of the ONS Management and Expectations Survey

Employer survey
• Ownership structure and networks • Management skills and locus of decision-making • Remote working • Use of targets and key performance indicators • Business improvement practices • Recruitment practices • Training and development • Performance management • Promotion practices and performance pay • Expectations about future business outcomes • Use of technology and AI

4.4 Illustrative uses

Two illustrative uses of linked, survey-based LEED are provided in Appendix B.

In **Error! Reference source not found.**, we outline the potential to use survey-based LEED to understand the factors that drive employee engagement and commitment in organisations, and the role that employee attitudes and behaviours play in driving firm performance. We expect this use of LEED would primarily be of interest to the Department for Business and Trade, Acas and HM Treasury.

In **Error! Reference source not found.**, we outline the potential to use survey-based LEED to understand the factors that shape job quality and employee well-being. We expect this use of LEED would primarily be of interest to the Department for Business and Trade, Acas, the Low Pay Commission and the Department for Work and Pensions.

Next steps

In the previous sections of this report, we have argued that the UK can – and should – build a new LEED infrastructure to provide timely and authoritative information on a range of policy-relevant issues such as business dynamism, productivity, wage growth and job quality. In contrast to many competitor countries, a comprehensive and high-quality LEED infrastructure does not currently exist in the UK. However, this could be developed by building a LEED spine, based on HMRC PAYE data, to which a range of existing administrative data and new surveys can be linked for research use.

The development of this LEED infrastructure is technically feasible, and there is a clear desire across government and academia for the UK to have it. The potential value is significant. If the infrastructure is well designed, government would have better information when designing policy, economic and social science research would be boosted by access to high-quality data, and one would avoid unnecessary duplication by replacing ad hoc efforts with a sustainable and coherent program of data development and data linking.

In our view, the roadmap for the development of this new LEED infrastructure comprises three related stages:

- **Stage 1:** Utilise the PAYE RTI data to develop a **LEED spine** which will provide unique and time-consistent identity linkages for employees and employers, allowing workers and firms to be tracked over time
- **Stage 2:** Develop **linkages** from the LEED spine to a range of existing administrative and survey data sources, so that additional data on employers and employees can be added to the spine
- **Stage 3:** Use the spine as a **sampling frame** for new employer-employee surveys, so that gaps in our existing measurement of the employer-employee relationship can be addressed.

To move forward, we propose that ESCoE should collaborate with ONS and HMRC over the next 12 months to develop a prototype for a new LEED infrastructure, consisting of a LEED spine linked to a small number of additional datasets. In doing so, ONS would build on its unique position as the UK's National Statistical Institute with unrivalled capabilities to create research-ready datasets through data linkage. It would also build on its experience with the PAYE RTI data and its experience in building the LBD as a reference spine of businesses. Fundamentally, it would also build on its existing relationships with HMRC, as data owner of the PAYE RTI data. ONS and HMRC have a critical role to play as data custodians, providing continuity, trust and credibility in the eyes of data owners, researchers and future survey participants. ESCoE would provide academic expertise in the creation and analysis of LEED. Additional collaborative partnerships should also be sought with others in government and academia who have experience of using PAYE RTI.

The goal, however, is to go beyond a prototype to create a sustainable LEED infrastructure that exists for many years to come. Thus, attention must be paid to developing the prototype into a fully-fledged infrastructure.

To achieve this goal, there will be an important role for government departments in supporting data linkage, and we envisage a key role for ADR UK in supporting these linkage projects under its next phase of funding, beginning in April 2026. We also envisage an important role for government and ESRC in supporting new surveys which use the LEED spine as a sampling frame. Policy makers and researchers will also be critical in arguing for the creation of the data infrastructure, shaping its development and ensuring that (if implemented) it is used to generate an evidence base that impacts policy-making.

We hope that all parties will seize the opportunity that we believe now presents itself over the next 2-3 years for the UK to build a new LEED infrastructure for the public good.

References

Abowd, J. M., Stephens, B., Vilhuber, V. Andersson, F., McKinney K., Roemer, M. and Woodcock, S. (2005). The LEHD Infrastructure Files and the creation of the Quarterly Workforce Indicators. Chapter 5 (pp. 149-230) in T. Dunne, J. Jensen and M. Roberts (eds.) *Producer Dynamics: New Evidence from Micro-Data*, University of Chicago Press.

Abowd, J. M., McKinney, K. L., & Zhao, N. L. (2018). Earnings inequality and mobility trends in the United States: Nationally representative estimates from longitudinally linked employer-employee data. *Journal of Labor Economics*, *36*(S1), S183-S300.

ADR UK (2025) *The Legal Framework for Using Administrative Data for Research Purposes*. Swindon: Administrative Data Research UK (ADR UK).

Bachmann, R., Bayer, C., Merkl, C., Seth, S., Stüber, H. and Wellschmied, F. (2021). Worker churn in the cross section and over time: New evidence from Germany, *Journal of Monetary Economics,* 117, 781–797.

Bagger, J., Fontaine, F., Postel-Vinay, F., & Robin, J. M. (2014). Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics. *American Economic Review*, *104*(6), 1551-1596.

Belzil, C. (2000). Job creation and job destruction, worker reallocation, and wages. *Journal of Labor Economics*, *18*(2), 183-203.

Bender, S., Bloom, N., Card, D., Van Reenen, J. and Wolter, S. (2018) Management practices, workforce selection, and productivity. *Journal of Labor Economics*, *36*(S1), pp.S371-S409.

Bloom, N., Ohlmacher, S. W., Tello-Trillo, C. J., & Wallskog, M. (2021). Pay, productivity and management, NBER Working Paper No. w29377.

Bobbio, E., & Bunzel, H. (2018). The Danish Matched Employer-Employee Data. Department of Economics Working Paper No. 2018-3, Aarhus University.

Buffington, C., Foster, L., Jarmin, R., and Ohlmacher, S. 2017. The Management and Organizational Practices Survey: An Overview. *Journal of Economic and Social Measurement*, 42:1–26.

Card, D., Heining, J., & Kline, P. (2013). Workplace heterogeneity and the rise of West German wage inequality. *The Quarterly Journal of Economics*, *128*(3), 967-1015.

Card, D., Rothstein, J. and Yi, M. (2024). Industry wage differentials: A firm-based approach. *Journal of Labor Economics*, 42, S1: S11-S59.

Card, D., Rothstein, J. and Yi, M. (2025). Location, location, location. *American Economic Journal: Applied Economics*, 17, 1: 297-336.

Casarico, A., & Lattanzio, S. (2024). What firms do: Gender inequality in linked employer-employee data. *Journal of Labor Economics*, *42*(2), 325-355.

Criscuolo, C., Gal, P., Leidecker, T. and Nicoletti, G. (2021) The Human Side of Productivity: Uncovering the Role of Skills and Diversity for Firm Productivity, OECD Productivity Working Paper No. 29, Paris: OECD.

Criscuolo, C., Haramboure, A., Hijzen, A., Koelle, M., & Schwellnus, C. (2022). The role of firm performance and wage-setting practices in wage inequality, Chapter 4 (pp.200-242) in OECD (ed.) *OECD Employment Outlook 2022: Building Back More Inclusive Labour Markets*, Paris: OECD Publishing.

Criscuolo, C., Hijzen, A., Schwellnus, C., Barth, E., Bertheau, A., Chen, W.H., Fabling, R., Fialho, P., Garita, J., Gorshkov, A. and Grabska-Romagosa, K. (2023a) Worker Skills or Firm Wage-Setting Practices? Decomposing Wage Inequality across 20 OECD Countries, Chapter 2 in OECD (ed) *The Role of Firms in Wage Inequality: Policy Lessons from a Large Scale Cross-Country Study*. Paris: OECD.

Criscuolo, C., et al (2023b). The firm-level link between productivity dispersion and wage inequality: A symptom of low job mobility?, Chapter 3 in OECD (ed) *The Role of Firms in Wage Inequality: Policy Lessons from a Large Scale Cross-Country Study*. Paris: OECD.

Davis, S. J., Faberman, R. J., & Haltiwanger, J. (2006). The flow approach to labor markets: new data sources and micro–macro links. *Journal of Economic Perspectives*, *20*(3), 3-26.

Dorsett, R. and Hug, J. (2022) Using UK tax records to produce new statistics on labour market transitions. ESCoE Discussion Paper No. 2022-10.

Drew, H., King, A. and Ritchie, F. (2013) *Impact Evaluation: Workplace Employment Relations Survey and European Social Survey. Final Report to the ESRC*. Bristol: University of the West of England.

Elsby, M., Shin, D. and Solon, G. (2016). Wage adjustment in the great recession and other downturns: Evidence from the United States and Great Britain. *Journal of Labor Economics*, 34, 1: 249-291.

Forth, J., Bryson, A., Phan, V., Ritchie, F., Singleton, C., Stokes, L. & Whittard, D. (2024). Revisiting Sample Bias in the UK's Annual Survey of Hours and Earnings, with Implications for Estimates of Low Pay and The Bite of the National Living Wage. ESCoE Discussion Paper No. 2024-10.

Fredericksen, A., Halliday, T. and Koch, A. (2016). Within- and cross-firm mobility and earnings growth. *Industrial and Labor Relations Review*, 69, 2: 320-353.

Fuest, C., Peichl, A., & Siegloch, S. (2018). Do higher corporate taxes reduce wages? Micro evidence from Germany. *American Economic Review*, *108*(2), 393-418.

Gomes, P. (2012) Labour market flows: facts from the United Kingdom. *Labour Economics*, 19: 165-175.

Graham, M., McEntarfer, E., McKinney, K., Tibbets, S., & Tucker, L. (2022). *LEHD Snapshot Documentation, Release S2021_R2022Q4* (No. 22-51). US Census Bureau.

Hahn, J., Hyatt, H., Janicki, H. and Tibbets, S. (2017). Job-to-job flows and earnings growth. *American Economic Review: Papers and Proceedings*, 107, 5: 358-363.

Haltiwanger, J., Jarmin, R. S., Kulick, R., & Miranda, J. (2017). High growth young firms: contribution to job, output, and productivity growth. In J. Haltiwanger, E. Hurst, J. Miranda and A. Schor (eds.) *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges* (pp. 11-62). University of Chicago Press.

Haltiwanger, J., Hyatt, H. and McEntarfer, E. (2018). Who moves up the job ladder?. *Journal of Labor Economics*, 36, S1: S301-S336.

Heining, J., Klosterhuber, W., & Seth, S. (2014). An overview on the linked employer-employee data of the Institute for Employment Research (IAB). *Journal of Contextual Economics–Schmollers Jahrbuch*, (1), 141-148.

Hennig, J.L. and Stadler, B., 2023. Firm-specific pay premiums and the gender wage gap in Europe. *Economica*. 90, 359: 911-936.

HMRC (2023). *Guidance on RTI Data Items from April 2024*, Version 1.1, dated 24th October 2023. London: HMRC.

HMRC (2025). Consultation outcome. Draft legislation: Improving the data HMRC collects from its customers. 28th January. London: HM Revenue and Customs.

Huffman, M. L., King, J., & Reichelt, M. (2017). Equality for whom? Organizational policies and the gender gap across the German earnings distribution. *ILR Review*, *70*(1), 16-41.

Huselid M (1995) "The impact of human resource management practices on turnover, productivity and corporate financial performance", *Academy of Management Journal*, 38, 3: 635-872.

Integrated Data Service (2025). *PAYE RTI IDS User Guide*. London: Integrated Data Service.

Lane, J., and Stephens, B. (2006). Integrated employer-employee data: New resources for regional data analysis. *International Regional Science Review*, *29*(3), 264-277.

Lawler E (1986). *High Involvement Management*, San Francisco: Jossey Bass.

Leaker, D. and Taylor, C. (2019). *New methods for monthly earnings and employment estimates from Pay As You Earn Real Time Information (PAYE RTI) data*. London: Office for National Statistics.

Leaker, D. (2022). *Comparison of Labour Market Data Sources*. London: Office for National Statistics.

Lemma, Y., Lui, S., Romaniuk, L., Schneebacher, J., & Wolf, N. (2023). The UK Longitudinal Business Database, ESCoE Technical Report TR-23.

Mackeben, J., Ruf, K., Wolter, S. and Grunau, P. (2023) LPP survey data linked with administrative data of the IAB (LPP-ADIAB) 1975– 2021. FDZ-DatenReport 11-2021.

United Nations (2025) Labour. Chapter 16 in *System of National Accounts, 2025 (draft)*. New York: United Nations Department of Economic and Social Affairs.

ONS (2022a). *Monthly earnings and employment estimates from Pay As You Earn Real Time Information (PAYE RTI) data: methods*. London: Office for National Statistics.

ONS (2022b). *Job changers and job stayers, understanding earnings, UK: April 2012 to April 2021*. London: Office for National Statistics.

ONS (2024a). *Business Structure Database – UK*, Released 18 March. DOI: 10.57906/7kh0-0910. London: Office for National Statistics.

ONS (2024b). *The impact of bariatric surgery on monthly employee pay and employee status, England: April 2014 to December 2022*. London: Office for National Statistics.

ONS (2024c). *Trends in UK business dynamism and productivity: 2024,* London: Office for National Statistics.

ONS (2025a). *Earnings and employment from Pay As You Earn Real Time Information, seasonally adjusted*. Release dated 20 March 2025. London: Office for National Statistics.

ONS (2025b). *X02: Labour Force Survey flows estimates*. Dataset. London: Office for National Statistics.

ONS (2025c). *The impact of an endometriosis diagnosis on monthly pay and employee status, England: April 2016 to December 2022*. London: Office for National Statistics.

ONS (2025d). *The Reference Data Management Framework: Data Architecture, Location and Integration*. London: Office for National Statistics.

Østergaard, C. R., Timmermans, B., and Kristinsson, K. (2011). Does a different view create something new? The effect of employee diversity on innovation. *Research policy*, *40*(3), 500-509.

Parrotta, P., Pozzoli, D., and Pytlikova, M. (2014a). The nexus between labor diversity and firm's innovation. *Journal of Population Economics*, *27*, 303-364.

Parrotta, P., Pozzoli, D., and Pytlikova, M. (2014b). Labor diversity and firm productivity. *European Economic Review*, *66*, 144-179.

Phan, V., Singleton, C., Bryson, A., Forth, J. and Ritchie, F., 2023. Accounting for firms in gender-ethnicity wage gaps throughout the earnings distribution. University of Reading Department of Economics Discussion Paper No. 2023-16.

Powell T (1995) "Total quality management as competitive advantage: a review and empirical study", *Strategic Management Journal*, 16, 1: 15-37.

Prime Minister's Office (2019) *The Queen's Speech 2019: Background Briefing Notes*, London: Prime Minister's Office.

Ritchie, F, Phan, V., Whittard, D., Bryson, A., Caden, A., Forth, J., Scarfe, R. and Singleton, C. (2024) *Creating the PAYE Panel*. Bristol: WED project.

Ritchie, F, Phan, V., Whittard, D., Bryson, A., Caden, A., Forth, J., Scarfe, R. and Singleton, C. (2025) *Annual Survey of Hours and Earnings linked to Pay-As-You-Earn and Self-Assessment Data – England, Scotland, and Wales. ADR Data Explained*. London: ADR England.

Tanaka, S., Warren, L. and Wiczer, D. (2023). Earnings growth, job flows and churn. *Journal of Monetary Economics*, 135: 86-98.

Timmermans, B. (2010). The Danish integrated database for labor market research: towards demystification for the English speaking audience. DRUID Working Paper No. 10-16. Danish Research Unit for Industrial Dynamics.

Upward, R., and Wright, P. W. (2019). Don't look down: the consequences of job loss in a flexible labour market. *Economica*, *86*(341), 166-200.

UK Government (2024) *Next Steps to Make Work Pay*, London: Department for Business and Trade.

Wallskog, M., Bloom, N., Ohlmacher, S., Tello-Trillo, C. (2024) Within-firm pay inequality and productivity. NBER Working Paper 32240.

Appendix A: LEED infrastructures in other countries

The US Longitudinal Employer-Household Dynamics (LEHD) dataset

In the United States, the *Longitudinal Employer-Household Dynamics (LEHD) dataset* comprises a quarterly database of linked employer-employee data covering all wage earners. The core data on jobs and earnings are derived mostly from the state unemployment insurance systems, which is shared with the US Census Bureau under a partnership arrangement. Employee demographic characteristics are matched in from the decennial population Census and the annual American Community Survey, whilst employer characteristics are matched in from the Quarterly Census of Employment and Wages (QCEW) and Census Bureau Longitudinal Business Database (CBLBD). Annual information on firm sales and productivity are matched in from the revenue-augmented version of the CBLBD (see Haltiwanger, 2017), whilst data on firms' management practices can be linked in from the Managerial and Organizational Practices (MOPS) survey (see Buffington, 2017). Abowd et al (2004) and Lane and Stephens (2006) provide general introductions to the dataset, whilst detailed metadata is presented in Graham et al (2022). Data is available from most states from the late 1990s onwards.

LEHD data is used to generate a range of indicators of labour market dynamics, including the Census Bureau's Quarterly Workforce Indicators (*QWI*), LEHD Origin-Destination Employment Statistics (*LODES*), Job-to-Job Flows (*J2J*), and the Post-Secondary Employment Outcomes (*PSEO*) data. The LEHD microdata is available to approved researchers via the Federal Statistical Research Data Center (FSRDC) network. The microdata have been used to publish research on a number of topics, including the links between earnings inequality and job mobility (Abowd et al, 2018) and the links between pay, productivity and management practices (Wallskog et al., 2024).

The Danish Integrated Database for Labor Market Research (IDA)

The Danish *Integrated Database for Labor Market Research (IDA)* is an annual database combining information on the universe of labour force participants and workplaces in Denmark. The IDA is compiled from a variety of registers maintained by Statistics Denmark. Data on annual hourly earnings is derived from the Salary Information Register (SOR), with unemployment status also observed from the Unemployment Statistics Register (CRAM). Employee characteristics are matched in from the Population Register (CPR) and Education Register (UDDA). Employer characteristics and annual information from firms' accounts can be linked to the IDA from various firm-level registers (e.g. FIRE, FIRM) held by Statistics Denmark. Researchers have also linked the IDA to firm-level surveys of innovation practices (DISKO). Timmermans (2010) and Bobbio and Bunzel (2018) provide introductions to the dataset. Data is available from 1980 onwards. Research using IDA microdata includes studies of the wage returns to education and work experience (Bagger et al, 2014), the effects of job creation/destruction on wages (Belzil, 2000), and the effects of employee diversity on firm innovation (Ostergaard et al, 2011; Parrotta et al, 2014a) and productivity (Parrotta et al, 2014b).

The German Linked Employer-Employee Data (LIAB)

The German *Linked Employer-Employee Data (LIAB)* combines individual employment biographies generated from the social security system with establishment data collected from administrative sources and an annual establishment panel survey. The component datasets are compiled by the Federal Employment Agency (BA)

at the Institute for Employment Research (IAB). The Individual Employment Biographies (IEB) database provides information on all employment spells at a given establishment within a calendar year, and the average daily wage, taken from notices of employment filed by the employer. Data on the employer comes primarily from the IAB Establishment Panel - an annual representative survey of around 16,000 establishments, sampled from the BA's establishment register. The survey includes questions on employment, capital investment, turnover and a limited number of management practices. The Establishment History Panel (BHP) then provides additional data on the gender, age, occupational status and qualifications of the workforce at the establishment, compiled from individual social security records. Employee data is available from 1975 onwards; the IAB Establishment Panel extends back to the mid-1990s. Heining (2014) provides an overview of the LIAB, with the microdata having been used to publish research on a number of topics, including the role of workplace heterogeneity in driving wage inequality (Card et al, 2013; Huffman et al, 2016) and the effects of corporate taxes on wages (Fuest et al, 2018). Since 2012, the IAB has also used the LIAB as the basis for additional, periodic linked employer-employee surveys which collect additional data on human resource practices, employee work attitudes and wellbeing (data similar to that previously collected in the British *Workplace Employment Relations Survey*) (Mackeben, 2023).

Appendix B: Illustrative use cases

Use Case 1: Generating new indicators of business dynamism

What is the policy issue to be addressed?

A dynamic economy has a healthy rate of job creation and destruction, facilitating the reallocation of labour from less productive to more productive enterprises. So a lack of business dynamism could lie behind stagnation in productivity or wages (Goldin et al., 2024; Hurst and Kahn, 2023).

The long-standing approach to measuring business dynamism has focused on changes in the stock of jobs in the firm (also known as "job flows"). Job creation is estimated by measuring net employment increase in growing firms and the stock of employment in new firms. Similarly, for job destruction. However, this approach captures only part of the picture. A business with two quits during a period and one new hire has a net change of one destroyed job, but this arises through the outflow of two workers and the inflow of a third. A more complete understanding requires data on "worker flows", so that one can measure enterprises' hiring activity and job separations, and the extent to which hires and separations arise through worker reallocation across firms or movements from/to non-employment.

Worker flows and job flows are necessarily related. However, worker flows provide a more sensitive measure of business dynamism because they capture within-firm reallocation (e.g. laying off workers in one area and hiring in another). They also better capture short employment spells, which may arise from uncertainty over match quality, or highly-volatile patterns of demand. Davis et al. (2012) provide an analysis of job flows and worker flows for the US using the LEHD. Bachmann et al. (2021) provide a similar analysis for Germany using a dataset derived from the LIAB.

Why can this issue not be addressed satisfactorily with existing data sources?

The UK's statistics on business dynamism are primarily generated using the LBD which provides measures of net change in total employment for each business. This yields statistics on rates of job creation, job destruction and job reallocation (see ONS, 2024c). However, the LBD does not contain data on worker flows, and so it is unable to decompose net changes in stocks into hiring and separations or identify extent to which hiring and separations arise from reallocation of workers across firms or movements from/to non-employment. Estimates of workers flows are available from the LFS (see ONS, 2025b; Gomes, 2012), but these lack granularity due to limited sample sizes, and they are not necessarily consistent with estimates from the LBD, being produced from a different data source.

How would the LEED spine fill this evidence gap?

A LEED spine would enable the identification of worker flows (and job flows) for each business. Specifically, it would enable the ONS to generate internally consistent measures of hires, separations, job creation, and job destruction at the enterprise and aggregate level. It would also enable the measurement of worker reallocation across firms and movements to/from non-employment on a basis that is consistent with measures of hiring and separations. Finally, it would provide estimates at a more granular level than is currently possible from the LFS, notably by detailed industry sector. This would provide a more complete understanding of business employment dynamics in the UK.

Use Case 2: Generating new indicators of wage growth

What is the policy issue to be addressed?

The UK has experienced slow growth in average earnings over the past decade. However, even when there is little change in the average rate of earnings growth, individual workers experience a distribution of earnings changes. Some achieve earnings growth within the same job as they gain experience, others grow their earnings by moving to higher-paying firms, and still others see their earnings fall as a result of job loss.

Headline measures of wage growth typically rely on unconnected firm and worker-level data. This provides a descriptive portrait of earnings changes, particularly at the mean. However, related to the issues outlined in **Error! Reference source not found.**, this provides only a limited understanding of the dynamics that underlay wage changes at the aggregate level. A fuller understanding can be obtained if one is able to observe the earnings changes associated with different worker flows in the labour market (i.e. separating the wages paid to new hires from wage growth on the job). This can aid our understanding of the influence of cyclical factors on wages, the degree of wage rigidity within ongoing employment contracts, and the implications of job loss, e.g. through decarbonisation.

Hahn et al. (2017) and Tanaka et al. (2023) use the LEHD to reveal how the contribution of different flows to aggregate wage growth changes in the US. Hijzen et al. (2024) present a comparative analysis of the US and Norway to understand the impact of different economic systems on wage dynamics. Abowd et al. (2018) go beyond the mean, using the analysis of worker flows to understand changes in earnings inequality in the US.

Why can this issue not be addressed satisfactorily with existing data sources?

Average weekly earnings (AWE) is the ONS' lead indicator of short-term changes in earnings. AWE is based on the Monthly Wages and Salaries Survey (MWSS), which collects aggregate data from businesses on the total amount paid and number of paid employees. It provides no worker-level data.

The Labour Force Survey (LFS) collects information on the earnings of individual employees, and in its longitudinal form, observes worker flows. However, the earnings data is not considered to be of high quality. ASHE is considered to provide more reliable earnings data, deriving from employers' payrolls (as in the case of MWSS), and has been used to analyse the dynamics underlying wage growth (see Elsby et al., 2016; ONS, 2022b). However, ASHE suffers from non-random panel attrition and it is not possible to separate panel exit due to job loss from exit due to employer non-response (see Forth et al., 2024).

How would the LEED spine fill this evidence gap?

A LEED spine would enable the identification of the earnings changes associated with worker flows into and out of employee jobs. This would enable the decomposition of growth at the mean (or at different points of the earnings distribution) into the contributions from: changes in hiring wages, earnings losses of displaced workers, earnings growth arising from job-to-job moves, and earnings growth among those who remain in the same job. Any analysis would necessarily be based on a calenderised measure of weekly or monthly earnings from PAYE RTI. This implies that analysis of ASHE would continue to be useful as a means of understanding the importance of changes in hours worked.

Use Case 3: Using data linked to the LEED spine to evaluate the impact of changes to National Insurance

What is the policy issue to be addressed?

The 2024 Autumn Budget announced changes to National Insurance (NI) contributions for employers. The changes included raising the contribution rate, lowering the threshold below which no NICs are due, and increasing the Employment Allowance. This complex set of changes will impact firms in different ways depending on the number of staff they have and the pay of those staff, though for most firms it will represent an increase in payroll costs. Firms are likely to adjust along one or more margins: raising prices, reducing wages (relative to otherwise), reducing headcount (by either firing or reducing hiring), reducing hours worked, investing in capital, accepting lower profit margins. These effects are each of interest in themselves. However, the combination of employer responses also has the potential to have meaningful implications for inflation and monetary policy.

Why can this issue not be addressed satisfactorily with existing data sources?

Current datasets do not provide enough detail about the firm and its workers to fully analyse the effects of the policy. The Annual Survey of Hours and Earning (ASHE) includes data on pay of employees, and contains some firm-level information; but as a 1% sample of employees across the whole economy, there is very little scope to analyse pay distributions *within* firms, which are important as a means of identifying businesses that are most (or least) affected by changes in NIC earnings thresholds. Other datasets like the Monthly Wages and Salaries Survey contain data on firm-level average wages, but not on the distribution of earnings across workers within the firm. MWSS also over-represents larger firms, while the impact will be proportionately greater for smaller firms. Without a link to firm level data on business performance or investment, it is also difficult to evaluate the impact of the NI changes on business decisions.

How would a new LEED infrastructure fill this evidence gap?

Comprehensive LEED based on PAYE RTI data would allow comprehensive assessment of the effect of the NICs changes on firms, since it would contain information on the pay of all workers within each firm alongside a range of worker and firm characteristics. Analysts could use these data to identify exposure to the policy, and then to examine some of the changes that occur within firms. If the LEED were regularly updated (e.g. monthly or quarterly), it could be used to monitor the effects of the current NICs changes. The LEED spine would offer some opportunities for analysis, in terms of observing changes in headcount, hiring rates, separation rates and labour costs. However, it would be particularly useful if the LEED could also have linked data from other surveys, such as data on vacancies from the Vacancies survey, data on business performance from the ABS survey, and on capital investment from the QCAS and/or ABS survey.

For a LEED to be of most use in this context, it would need to be accessible within the next year, and ideally as soon as possible.

Use Case 4: Using data linked to the LEED spine to analyse productivity growth and wage growth

What is the policy issue to be addressed?

Productivity growth has been weak in the UK for over a decade. Real wage growth has also been slow for much of that period. Trends in productivity and wages are of major interest to policy makers: productivity is a key parameter determining the growth rate of the economy, and wage growth (especially over and above any improvements in labour productivity) has important implications for inflation and monetary policy. However, many things about wage-setting and productivity growth in the UK are not known due to a lack of suitable data sources. In particular, it would be valuable to better understand the extent to which wages and productivity are driven by firms or workers (or their interaction). Some research questions include:

- Do more productive firms attract more productive workers?
- How does the nature of worker-firm sorting vary across sectors and regions, and to what extent can it explain the diffusion of innovation, and trends in productivity and wages?
- How much do worker characteristics matter for workers' earnings, compared to firm-level productivity and pay setting?
- Is recent pay growth driven by an improvement in worker skills, the result of job switching, or due to changes in how firms set pay?

Analysis of such questions is commonplace in countries with LEED data. Examples from the US and Germany include: Bender et al. (2018); Haltiwanger et al. (2018); Wallskog et al. (2024); and Card et al. (2024, 2025).

Why can this issue not be addressed satisfactorily with existing data sources?

The decomposition of wages and productivity into firm, worker and match components is usually done by estimating two-way fixed effects (AKM) models. To estimate such a model, one needs to observe multiple workers per firm and a sufficient number of workers moving across firms, alongside worker and firm characteristics. Existing data sources (such as ASHE) are too sparse, including only a very small number of workers employed at the same firm, and making it near impossible to track workers moving between firms. Similarly, there are no UK datasets which give a comprehensive view of productivity dynamics from the perspective of both the firm and the worker (e.g. offering data on firm productivity, management practices, labour quality and so on).

How would a new LEED infrastructure fill this evidence gap?

Comprehensive PAYE RTI data linked to data on businesses (from the LBD and ABS) would create a sufficiently rich dataset allowing for the decomposition of earnings and productivity into firm, worker and match components. This would allow us to answer some of the questions above for the UK.

A broader data infrastructure linking LEED to survey data would allow for an even richer analysis of pay and productivity, such as by adding information on detailed worker characteristics (e.g. qualifications data from the Census or HESA) and firms' use of technology and management practices (e.g. from the MES). Academic literature on productivity and pay, based on LEED in other countries, has shown the importance of management practices and worker sorting for productivity (e.g. Bender et al., 2018), but we have little empirical evidence on whether and how much they matter in the UK.

Use Case 5: Using data linked to the LEED spine to identify risk factors associated with economic inactivity

What is the policy issue to be addressed?

The Get Britain Working White paper outlines plans to get more people *into* work and to enable more people to *get on at* work. A primary focus is on reducing inactivity by boosting employment and boosting progression. Existing evidence on inactivity shows that there are important links between health and one's ability to work. It is also widely recognised that the work setting can be important in helping people to remain in work. There are also significant challenges in the youth labour market, particularly in the transition to employment, further education, and training. The rising number of young people not in education, employment, or training (NEET) underlines the importance of addressing this issue.

Improving transitions into work—and from education into employment or training—is essential. Equally important is reducing exits from work into inactivity. To better target policy interventions, it is crucial to understand labour market dynamics and the role that individual and employer characteristics play in these. One consequence of an improved understanding of these factors is the development of 'risk of economic inactivity' indicators paralleling in some ways the 'risk of NEET indicators' that already exist and can be used to better target policy interventions on those at greatest risk of becoming NEET.

Why can this issue not be addressed satisfactorily with existing data sources?

Setting ongoing data quality issues aside, LFS/APS provide rich information on an individual's personal characteristics and economic activity status (covering all states). It also provides information on their job characteristics, but only limited information about their work setting, and has a limited follow-up period (one year). ASHE provides good-quality information on an employee's earnings and hours, with the potential to follow individuals over many years. The data can also be linked to ONS business data (e.g. on management practices). However, ASHE contains few personal characteristics, has considerable panel attrition and provides no information on a person's status if they cease to be an employee.

LEO offers comprehensive longitudinal data on an individual's personal characteristics, especially their educational attainment, and some information about their employment. However, in common with the LFS, it provides only limited information about the work setting. Coverage is also mostly limited to people born since 1985. The DWP's RAPID dataset covers all persons with a National Insurance number. It provides annualised and point-in-time information about their employment status, earnings and benefit receipt. However, it does not (yet) include information about educational attainment and provides very little information on the work setting.

How would a new LEED infrastructure fill this evidence gap?

A new LEED infrastructure could provide longitudinal information on pay and progression in the workplace (from PAYE RTI), linked to information on benefit receipt and non-work episodes (from RAPID), health data (hospital episodes), demographic characteristics (from the Census), education (from HESA) and firm management practices (from MES). This would extend our understanding of the risk factors for economic inactivity, as well as the prospects for in-work progression by providing new insights into the role that individual and employer characteristics play in individuals' labour market trajectories.

Use Case 6: Using data linked to the LEED spine to understand the "winners" and "losers" from trade

What is the policy issue to be addressed?

The UK's decision to withdraw from the EU has intensified demand for higher quality statistics, including on trade and employment. For the first time in 50 years, the UK has to set its own trade policy, and this means there is increased demand on the evidence to support such policies. The Centre for Inclusive Trade Policy (CITP) citizens juries research demonstrates that the public has a complex approach to weighing up trade-offs that can arise from trade, and this has implications for expectations on government policy. For example, potentially prioritising regional fairness above maximising national growth. More research is needed to investigate the "winners" and "losers" from trade, helping us understand what types of firms export, their impact on the economy, how many people they employ and the characteristics of those they employ. Illustrative research questions include:

- Do the employees of trading firms differ from those of non-trading firms in terms of skills, regional location and other characteristics (e.g. gender, age, ethnicity, disability)?
- Do employees in trading firms enjoy a wage premium? If so, what are the origins of this?
- How does this link into supply chains across the UK?

Why can this issue not be addressed satisfactorily with existing data sources?

Existing research has been able to identify trading firms by linking trade data to firm-level data. However, to say something precise on the characteristics of people employed by trading firms requires linked data on employees. This firm-to-person link is challenging with currently available data. ASHE has a limited sample size and only provides characteristics for location/age/sex. This means that much of the existing work is based on industry averages.

How would a new LEED infrastructure fill this evidence gap?

Using the LEED spine to link trade data and firm-level data to individual data on employees would enable new research to account for a much wider set of employee characteristics than hitherto (including education, ethnicity, deprivation etc.). The large sample sizes in PAYE RTI would also address many sample size challenges. Such research would improve our ability to assess the impact of potential future trade agreements and support trade negotiations. It would also help develop economic strategies which help all people achieve their full economic potential.

Use Case 7: Using new surveys sampled from the LEED spine to understand the human side of productivity

What is the policy issue to be addressed?

There are large differences in productivity between firms. Some of these differences are related to differences in workforce skills and use of structured management practices (see **Error! Reference source not found.**). But the impact of skills and management practices on productivity is mediated through employee decision-making and effort. This implies an important role for understanding the factors that drive employee engagement and commitment in organisations, and the role that employee attitudes and behaviours play in driving firm performance. It also indicates the importance of understanding employees' role in promoting or inhibiting organisational change. Some research questions include:

- Why are productivity-enhancing management practices not more widely adopted throughout the economy? Are there gaps in the incentives that are needed for managers to be able to configure and implement such practices, and for employees to work effectively within such systems?
- To what extent is productivity growth dependent on the nature of the employment relationship within the firm? How important are employee engagement, job satisfaction and employee wellbeing in enabling firms to maximise the benefit of structured management practices?
- What role do trade unions and other forms of employee involvement and voice have in promoting productivity growth, innovation and organisational transformation?

Why can this issue not be addressed satisfactorily with existing data sources?

To answer these questions requires rich, linked information from both the employer and employee side of the employment relationship. WERS provided rich information on how employees were managed in the workplace, and on employee attitudes to work. However, the last survey was conducted in 2011, the data were largely cross-sectional and the survey provided limited data on firm performance. The MES provides data on structured management practices but has no employee component. Employee surveys, such as the Skills and Employment Survey and CIPD's UK Working Lives Survey provide data on employees' attitudes to work but have no employer component.

How would a new LEED infrastructure fill this evidence gap?

Using new surveys to collect data on workforce skills and employee attitudes to work, and linking these to data on management practices and firm performance, would provide new opportunities to investigate the diffusion of managerial innovations, and the role that employees' attitudes and behavioural reactions play in the 'black box' that links management practices to firm performance.

A new linked employer-employee survey that includes a broad conception of management practices, alongside other aspects of the firm's employment system and accounting data on performance, would also allow a better understanding of the role of trade unions and other forms of worker voice in promoting product and process innovation and limiting the adverse consequences of workplace innovation on employee wellbeing.

Use Case 8: Using new surveys sampled from the LEED spine to understand differences in job quality and employee well-being

What is the policy issue to be addressed?

Job quality has long been the subject of research and policy making. The previous government stated its ambition for the UK to become "the best place in the world to work" (Prime Minister's Office, 2019: 42), and the current Government has also put job quality at the heart of its policy agenda. They have stated their commitment to creating jobs that provide security, treat workers fairly, and pay a decent wage (UK Government, 2024). They have also stated their intention to make flexible working the default, to strengthen leave provisions and to promote a new approach to employment relations based on co-operation and negotiation. Such policy statements recognise that job quality is a key determinant of citizens' quality of life and overall well-being.

Job quality has typically been attributed to factors such as technological change, the supply of skills and macro-social trends. However, plentiful evidence now demonstrates that where you work has an impact on wages beyond these other factors. Much less is known about the importance of the firm in accounting for other dimensions of job quality. It is important to be able to investigate the role of the firm and other factors in shaping job quality, because this supports an understanding of the extent to which job quality (and in turn, worker well-being) can be shaped by regulatory controls. It is also important to be able to study the impact of policy initiatives in this area.

Why can this issue not be addressed satisfactorily with existing data sources?

Trends in job quality and well-being can be observed from surveys of employees, such as the Labour Force Survey, the Skills and Employment Survey and the CIPD Working Lives Survey. However, a fuller understanding of the determinants of job quality rely on linked employer-employee data that enables the employment relationship to be observed in the round. None of these surveys includes data from employers. As noted in **Error! Reference source not found.**, WERS captured rich information on job quality, job satisfaction and employee well-being alongside plentiful information about the employee's job and workplace setting. But these data are now dated.

How would the LEED spine fill this evidence gap?

New nationally-representative survey-based LEED could offer objective and subjective measures of job quality *alongside* measures of the employment practices and management strategies. It could also enable a more complete understanding of differences in employees' experience of work (the 'winners' and 'losers') by identifying the differing experiences of different groups of employees within the firm (e.g. men and women) and by identify the impact of employer strategies or characteristics that are difficult to observe in employee-only data. The value of LEED is further amplified when it offers measures of firm or workplace performance, because variations in productivity and firms' survival prospects have clear implications for job quality and employee well-being.