

# Coherence, Vacuity and Linkage

---

ASSESSING THE POTENTIAL TO USE GOVERNMENT ADMINISTRATIVE DATA TO IMPROVE  
SOCIAL AND ECONOMIC STATISTICS

MIRELA NIKOLOVA, EMMA WHITE

## What are Official Statistics and why are they produced?

---

All those statistical outputs produced by the UK Statistics Authority's executive office (the Office for National Statistics), by central Government departments and agencies, by the Devolved Administrations in Northern Ireland, Scotland and Wales, and by other Crown bodies.

*Statistics and Registration Service Act 2007*

- Inform government ministers, policy makers and political assemblies about the state of the nation and the economy and changes over time;
- Provide businesses with the information to help them run effectively and efficiently;
- Meet the needs of the European Union and other international bodies so people can compare data across countries;
- Help analysts, researchers, scholars and students with their work.

*Office for National Statistics, [www.ons.gov.uk](http://www.ons.gov.uk)*

At the moment...

---

Over 1000 National and Official statistical products are produced by 25 Government Departments and agencies and the three Devolved Administrations.

Examples of ONS statistics:

- Labour market statistics
- Migration statistics
- Overseas travel and tourism
- Families and households in the UK
- Business investment in the UK
- Business activity, size and location in the UK

UK Statistics Authority, [www.statisticsauthority.gov.uk](http://www.statisticsauthority.gov.uk)

## Assessing the potential

---

In December 2016, ONS commissioned 3 projects (running in parallel) to investigate the potential for several government administrative data sources to be used for producing Official Statistics.

- **Coherence** – to examine the definitions of key concepts within the administrative datasets and find out how they align with concepts used in primary government surveys.
- **Vacuity** – to investigate issues relating to under-coverage and partial coverage that are specific to each dataset.
- **Linkage** – to develop understanding of how key sources of information relating to individuals and businesses can be linked across the datasets.

# Data Sources

## Administrative datasets of interest

- HMRC
  - PAYE – Real-time information (PAYE-RTI)
  - Self-Assessment
  - Corporation tax
- DWP
  - Customer Information System (CIS)
  - Migrant worker scan (MWS)
  - National benefits database (NBD)
- NHS Digital
  - Personal Demographics System (PDS)
- Border Agency
  - Semaphore

## Surveys of interest

- Labour force survey (LFS) - employment, income from main/second job, work patterns, overtime, benefit entitlement, self-employment, home workers, health, etc.
- Annual survey of hours and earnings (ASHE) – pay and hours worked for employee in a business/company, including overtime pay, shift premium pay, bonuses, etc. (information provided by representative of the employer)
- Business register and employment survey (BRES) – number of employees, employment estimates, directors/shareholders
- Monthly business survey (MBS) – turnover, new orders, employment information from businesses

## Our process

---

### 1. Desk review of publically available sources

- ONS Sources (e.g., Census Transformation Programme)
- Official government documentation

### 2. Design of information-gathering templates and interview questions

### 3. Interviews and consultations

- Analysts from government departments;
- ONS representatives experienced in working with the data for various projects;
- Experts on different related topics (e.g., data linkage).

### 4. Consolidation of collected information and final project outputs

## Example: Pay As You Earn – Real-time Information

---

The Real-time Information system (RTI) was introduced in 2012 to streamline the PAYE process. It was piloted in two stages, with an increasing number of employers joining in phases. By October 2013, approximately 2.4 million payroll scheme operators joined.

- Employers make a full payment submission online each time they pay an employee;
- Approximately 65 million individual PAYE returns are made per month;
- RTI payment submissions can cover various pay periods (e.g., week, month year, irregular, etc.);
- There are 149 PAYE-RTI variables. Some are mandatory for every payment submission, some are optional and some only need to be completed once (e.g., when employee starts a new job).

## A matter of definition

---

Official government statistics on employment produced use the standard, international definition

*Persons in employment are defined as all those of working age who, during a short reference period, were engaged in any activity to produce goods or provide services for pay or profit. They comprise employed persons "at work", i.e. who worked in a job for at least one hour within a short reference period; and employed persons "not at work", due to temporary absence from a job (e.g., holiday, sickness, parental leave), or due to working-time arrangements (such as shift work, flexitime and compensatory leave for overtime).*

*International Labour Organization*

Is “employment”, as defined within administrative data, the same as “employment” according to the ILO?

## A matter of definition

---

Official government statistics on employment produced use the standard, international definition

*Persons in employment are defined as all those of **working age** who, during a short **reference period**, were engaged in any activity to produce goods or provide services for **pay or profit**. They comprise employed persons "at work", i.e. who worked in a job for **at least one hour** within a short reference period; and employed persons "not at work", due to **temporary absence** from a job (e.g., holiday, sickness, parental leave), or due to working-time arrangements (such as shift work, flexitime and compensatory leave for overtime).*

*International Labour Organization*

Is “employment”, as defined within administrative data, the same as “employment” according to the ILO?

## Reconciliation

---

Who is considered in “employment” in PAYE-RTI?

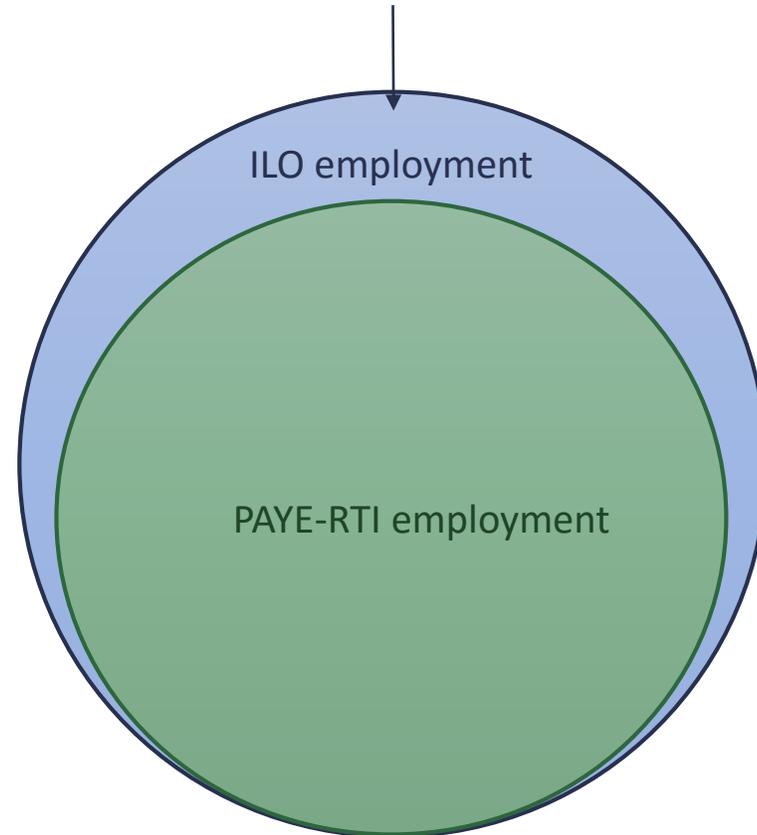
- Most individuals in full-time, part-time, irregular employment, currently on parental leave, apprenticeship/training scheme, 0-hours contract, etc.;
- Workers from abroad seconded to UK;
- Individuals with irregular payment pattern (e.g., paid to do one piece of work; seasonal workers, etc.);
- Students in full-time education working alongside their studies;

So far: Good alignment between PAYE employment and ILO employment!

Do the definitions overlap?

---

Self-employed (using Self-Assessment for tax returns)



## Vacuity

---

What are the main non-coverage/under-coverage issues in PAYE-RTI

- Employers who do not use the online submission system (very few);
- Employees in cases when everyone on payroll is paid under tax/NI threshold;
- Employees not paid via payroll (cash-in-hand);
- Volunteers;
- Unpaid family workers;
- Working illegally.

How can vacuity issues be addressed?

- Potential linkage with other data sources (e.g. Self-Assessment);
- Enhancing administrative data coverage with specific survey questions.

## Linkage

---

Linking records across different data sources can potentially produce very powerful datasets for analysis.

The challenges related to combining PAYE-RTI data with other datasets are nontrivial:

- Legal issues and ease of access;
- Differences in data quality standards, data entry procedures;
- Absence of common unique identifier across datasets;
- Update frequencies and accuracy of potential linkage variables;
- Ethical issues and re-identification concerns.

## Implications of this work

---

Potentially:

- With a sample of over 45 million employments each month, it may be possible to produce highly detailed, reliable data series on employment and income based on PAYE-RTI.
- Estimates of changes over time can be created at different levels of industry, geography and demography.
- Potential for using PAYE-RTI in the sampling design of social surveys, or as replacements for some questions on income and employment.

## Summary and conclusions

---

- Brought together administrative and survey metadata across different data sources
- Investigated Coherence, Vacuity and Linkage

Considering all datasets together, it may be possible to enhance or in some cases replace survey questions.

Using administrative data to its full potential would require in-depth knowledge and understanding of each dataset and collaboration with data-owning departments and analysts familiar with the data.

Next steps:

Recommendation to ONS about the potential of administrative data to enhance surveys and develop new questions.

Thank you!

---

## **Project team**

Emma White (University of Southampton, NatCen)

Paul Smith (University of Southampton)

Andy Cullis (University of Southampton)

Kirby Swales (NatCen)

Richard Boreham (NatCen)

Tricia Dodd (NatCen)

**NatCen**  
**Social Research** that works for society

UNIVERSITY OF  
**Southampton**