

	<p>Methodology Statement for Department for Work and Pensions Lifetime Labour Market Database publications</p>

January 2013

# Contents

- Introduction..... 3
- Tabulation Tool..... 4
  - Introduction..... 4
  - Loading - Stage 1 ..... 4
  - Production of the L2 Database - Stage 2..... 4
  - Quality Assurance - Stage 3..... 6
  - Pre-Publication ..... 6
  - Publication..... 7
- The DWP L2 First Release ..... 8
  - Introduction..... 8
  - Production ..... 8
  - Quality Assurance ..... 8
  - Pre-Publication ..... 9
  - Publication..... 9
- Annex A – Description of the Lifetime Labour Market Database (L2) ..... 10

# Introduction

The Lifetime Labour Market Database (L2) publications are published annually and aim to give users a structured overview of National Insurance Contributions and Qualifying years as well as Second Tier Pension Provision.

Alongside the First Release, there are a number of tables which enable the user to delve in to the detail:

- The DWP Tabulation Tool - for National Insurance Contributions and Qualifying years as well as Second Tier Pension Provision with an interactive tool to select one of many possible tabulations. Some statistics remain outside the scope of the Tabulation Tool but will continue to be available via the internet as pre-defined summary tables.
- Other tables and background information via links on the Tabulation Tool pages

The main purpose of this document is to provide users with information about all the methods used to produce the L2 publications in accordance with practices set out in the Code of Practice for Official Statistics.

# Tabulation Tool

## Introduction

The Tabulation Tool for the L2 publications provides the user with an interactive tool allowing users to select one of many possible tabulations (<http://83.244.183.180/NIRS/live/tabtool.html>). There are two main breakdowns of information to select from:

- National Insurance Contributions and Qualifying Years
- Second Tier Pension Provision

The information on the HMRC system is collected for calculation of National Insurance contributions, tax / refunds etc. This data was provided to DWP by HMRC initially to reformat and create files which could be analysed to produce information for the “Blue Book” and National Accounts. Since then, DWP have used the data to produce National Statistics and to provide analytical datasets for HMRC and GAD as well as for internal analysis across DWP for modelling, forecasting and analysis. Access is available internally for anyone with a relevant Business Case.

The Tabulation Tool is released annually. This publication previously had a February release until two years ago when DWP improved timeliness to publish in the previous December. In 2011, DWP have made it even earlier by releasing in October. In 2012, earlier still, publishing in April 2012.

## Loading - Stage 1

Production begins when DWP receive the 1% sample extract from HMRC. 65 data tables are received of which 14 are loaded into an Oracle database using a front end program called ASSIST.

Once the extracts have been loaded into ASSIST they are processed using pre-defined layouts to create SAS views. These SAS views form the base data for the 1% sample L2 Database and enable them to be interrogated in the same way as a SAS dataset, enabling the use of SAS in the Dataset and Cube creation for the Tabulation Tool and underlying datasets.

## Production of the L2 Database - Stage 2

Once the SAS views have been created, the L2 database, which forms the underlying datasets for the Tabulation Tool, is created. Using the most up to date methodology the L2 database contains 31 datasets. Of these 9 are used to directly feed in to the Tabulation Tool. Each dataset is created in a specific order and information from both raw data and datasets earlier in the chain are used to create subsequent datasets.

## Quality Assurance – Stage 2

Currently there is a QA program which produces frequencies and counts for certain key variables in each dataset, these figures are input into a spreadsheet that has all previous release QA figures in order for us to identify and plot changes in the data. All potential issues and suspect movements in the datasets are raised and investigated. It is often the case where movements can be attributed to change in National Insurance and pension policy and are fully expected.

Future development will include automated quality assurance carried out in SAS with manual analysis of output. The automated program aims to check for various features within the datasets compared to the previous version, such as missing variables, new decodes, missing values, and changes to the distribution of variables.

## Production of the Tabulation Tool – Stage 3

Once the quality assurance of the underlying datasets in the L2 database is completed and fully signed off, a process known as Cube creation begins. This part of the process is also carried out in SAS and re-formats selected SAS datasets into a format compatible with the next step, which is the Tabulation Tool creation process. The cubes set-up most of the desired categories for the Tabulation Tool, for example age variables are formatted into age bands and pension scheme codes are allocated to a category. There is one Cube for each of the two Tabulation Tools. Each Cube contains data from 1997 and all previously published data is updated to reflect the latest extract from HMRC.

Before the Tabulation Tool can be run, the Tabulation Tool control spreadsheets are updated. These spreadsheets contain every single parameter required for the run. All formats, titles, labels, links, and footnotes are defined here as well as which parts of the Tabulation Tool are required from the run. Changes to the spreadsheets are required each year. Changes in National Insurance and pension policy can often require additional footnotes. Once the changes are completed, the Tabulation Tool production can begin.

The Tabulation Tool is essentially a series of html pages which contain different cross-tabulations of data alongside appropriate titles, labels, and footnotes. JavaScript code creates the “front-end” parts of the Tool. This forms the interactive menus which sit on top of the html pages. The user makes selections on these pages and navigates to the appropriate table of data according to the selection made.

## Quality Assurance - Stage 3

Once the Tabulation Tool has been successfully created, the team quality assures the output. There are two questions to answer during this stage: Are the html pages correct and display all the correct data as expected? Does the Tool have all the correct functionality? The first question is addressed by use of an automated SAS checking program which directs attention to those pages showing the most change since the previous release. The idea behind this program is to compare the html page from the new edition with the identical html page from the previous edition. The program checks for things such as missing rows and columns from the table, missing footnotes, new footnotes and changed titles. It also assesses the level of change within the figures of the tables and flags those tables with the largest amount of change. All results are combined in an html report which is then checked by two members of the team. Anything listed in the html report is investigated to see whether it is expected or a potential issue. The second question requires members of the team to use and check the Tabulation Tool to ensure that all tables/ breakdowns/ footnotes are available and as expected. If an error is spotted, an incident will be raised and will be thoroughly investigated. If required, any part of the process can be re-run and all checking will re-occur.

Once completed, the lead Statistician will ensure everything has been checked and passed. Once happy with everything, they will sign off the Tabulation Tool. All actions are recorded on a release timetable, and are checked off once complete to ensure that no steps are missed.

## Pre-Publication

Once the Tabulation Tool has been signed off, it is then sent to the external web host in preparation for publication. The external web host currently hosts all of DWP's Tabulation Tools, including the L2 Tabulation Tool. All data from the previous release resides on the external site and is already live on the internet. When new data is created each year it must be uploaded onto the hosting site and replace the existing data.

When the data has been received by the external web host and uploaded for publication, it is the responsibility of the team to ensure that the same data exists on the hosting site as was sent. So, further quality assurance must be carried out at this stage. The external web host organises a secured pre-release window to the site for a few days in the week prior to release day and notifies the team of how to access this. A member of the team will then run an automated SAS checking program, which compares the Tabulation Tool on our system to that on the external web site. Provided there are no missing tabulations, no changes and no problems then we sign off the content on the external website and the pre-release access window is closed.

## Publication

The L2 Tabulation Tool is added to the external website at 09:30am on Release Day. Users can access the website via the DWP Internet pages. The L2 Tabulation Tool is accessible via the Create Your Own Statistics landing page which itself is available via the National Insurance Contributions and Qualifying Years and Second Tier Pension Provision Landing Page. The National Insurance Contributions and Qualifying Years and Second Tier Pension Provision landing page can be accessed here:

<https://www.gov.uk/government/organisations/department-for-work-pensions/series/national-insurance-contributions-statistics>

The Create Your Own Statistics landing page can be accessed here:

<https://www.gov.uk/government/organisations/department-for-work-pensions/series/dwp-statistics-tabulation-tool>

# The DWP L2 First Release

## Introduction

The L2 First Release brings together all of the relevant information and figures from the two aspects of the L2 publications and where necessary provides links to more detailed information. It can only be completed once all other products associated with the Tabulation Tool have been completed and quality assured. The First Release is a word document published on the DWP website as a pdf document. The First Release primarily contains headline figures (e.g., totals, top-level figures) and graphs. It is published annually.

## Production

The L2 First Release is a relatively simple document to produce. The format rarely changes between releases unless a major development needs addressing (in accordance with the Code of Practice); therefore production involves sourcing the relevant updated figures from the new Tabulation Tools.

All language, key points, and themes within the First Release have been kept consistent. Footnotes and notes are reviewed on a release by release basis to provide advice on the interpretation of the figures within the First Release.

The document contains a “Notes” section to reference other National Statistics products plus inform users of upcoming changes and issues.

## Quality Assurance

A member of the team takes responsibility for updating the First Release and once completed, it will be quality assured by a different member of the team. Production of the First Release involves sourcing figures from the newly produced Tabulation Tools and doesn't begin until each of these have been quality assured and signed off. Consequently, the team can be relatively certain the figures are correct at point of source. As the figures are sourced manually they still need to be quality assured in case of manual error when copying them across to the First Release. Once all figures have been quality assured they are then sent onto the lead Statistician for further checking and sign off. All actions are recorded on a release timetable, and are checked off once complete to ensure that no steps are missed.



## Pre-Publication

This is the most important part of the L2 publication process as this is pre-announced 12 months in advance and the deadline must be met, in accordance with release practices set out in the Code of Practice for Official Statistics. Dates of future publications can be found on the DWP internet via the statistics publication schedule web page (<https://www.gov.uk/government/organisations/department-for-work-pensions/about/statistics>) and via the National Statistics Publication Hub (<http://www.statistics.gov.uk/hub/index.html>)

As the L2 publications are National Statistics, there is an internal departmental requirement to accompany this release with a Submission. The Submission is circulated to the Minister's office and contains input from policy areas. The Submission is sent one day prior to release day and is fully compliant with the UKSA rules surrounding pre-release access of statistics. It contains a blend of the most notable issues within the latest release, key features, and policy guidance around these things. This document is also quality assured in the same manor as the First Release, before being sent out.

## Publication

The L2 publications are released on the DWP website at 09:30am on Release Day. The publication has its own landing page which contains the First Release as well as key information about the contents of the release and a link to the Tabulation Tool. The landing page can be accessed here:  
<https://www.gov.uk/government/organisations/department-for-work-pensions/series/national-insurance-contributions-statistics>.

# Annex A – Description of the Lifetime Labour Market Database (L2)

The L2 is a database system set up within the Department for Work and Pensions (DWP). It is based on a 1% download of Her Majesty's Revenue and Customs (HMRC) National Insurance Recording System (NIRS2) data which it cleans, validates and processes.

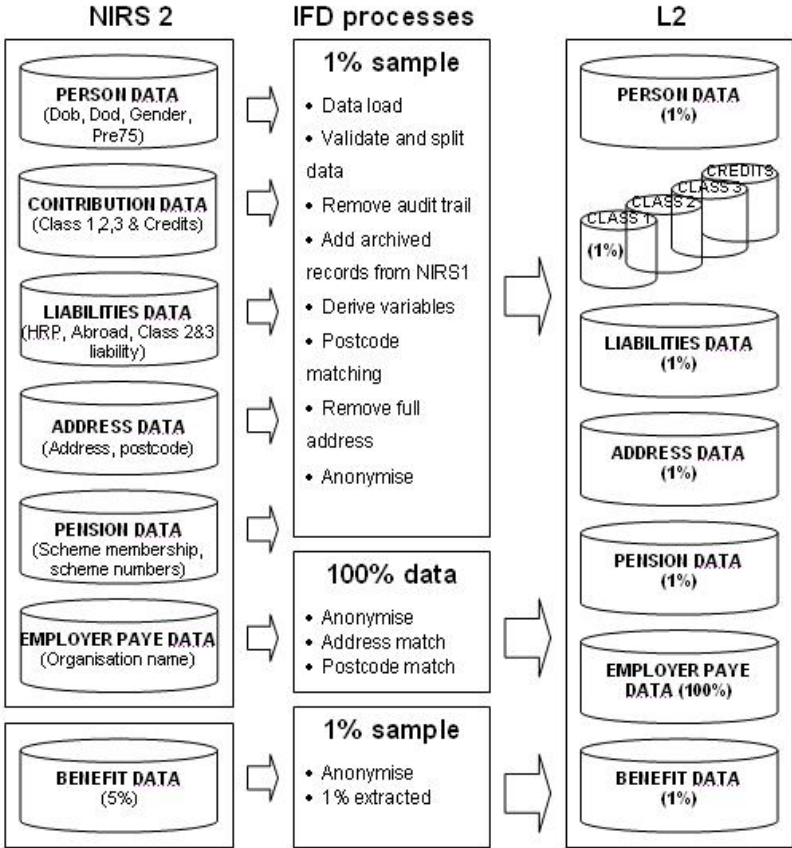
NIRS2 is the computer system for DWP and HMRC which contains all the information needed in the administration of National Insurance, Pay As You Earn (PAYE), benefits and pensions. It contains some demographic information and details of interactions of the population with DWP & HMRC going back to 1948 (in summary form) and going back to 1975 in detail. Some diagrams below give a basic pictorial representation of NIRS2 and how the L2 is created from it.

## Person data

NIRS2 holds National Insurance Numbers for all people aged 15¾ and over, who have either been allocated a National Insurance Number because they are a UK National, or who have arrived from abroad and have registered for a National Insurance Number for employment or benefit purposes. NIRS2 currently holds over 70 million National Insurance Numbers, including those for people who are no longer alive or who have left the country, meaning that longitudinal analysis of population and interactions are possible. An archiving exercise was performed in 1997 when data was transferred from NIRS1 to NIRS2 when only cases meeting certain criteria were archived. For the purposes of L2 we are able to 'recreate' these archived records from old extracts of NIRS1, although these cases would not be available from any new extract of NIRS2. The archiving exercise means that a new NIRS2 will be deficient of records prior to 1984, (primarily older women).

The person data holds information about date of registration for National Insurance purposes, date of birth, date of death, and information about National Insurance entitlements accrued before 1975 (when National Insurance records were computerised).

**Methodology Statement for the Department for Work and Pensions L2 publications**



**Contribution data**

The raw data from NIRS2 holds information about Class1 (employee), Class2 (self employment), Class3 (voluntary) and National Insurance Credits (contributory benefits) contributions in two data tables. The majority of the information is held in data table INPUT\_HEADER, which identifies the type of contribution, (Class1, Class2 etc). A single INPUT\_HEADER record will exist for each employment, class2 payment, voluntary contribution or National Insurance benefit credit in each tax year. For example, if there were two employments in the year, one payment of 52 weeks Class 2, as well as two spells on contributory benefit (JSA for example), there would be five entries on the INPUT\_HEADER table, each one being identifiable separately.

Underneath INPUT\_HEADER sits the data table CONTRIBUTION\_ITEM, which holds the detailed information about the category of contribution for Class 1 contributions only. For example, if a person paid two categories of National Insurance contribution in the year (Contracted Out (Category D) and Contracted In (Category A)), there would be two entries under a single INPUT\_HEADER record.

As part of the L2 creation process these data tables are linked to enable a comprehensive analysis of employment spells.

## **Methodology Statement for the Department for Work and Pensions L2 publications**

The raw NIRS2 data also includes an audit trail of all contribution items. If a contribution is amended, or removed, the audit trail marks the deleted/amended contribution and creates a new entry holding the corrected details. As part of the L2 creation processes the audit trail is removed, allowing only the corrected information to be passed forward.

All the data on contributions is tax year based. Detailed tax year information is available from 1975 onwards. The information will be recorded as follows:

- Class 1 – A record for each employment in the tax year. There may also be up to four contributions per employment record.
- Class 2 – A record for each Class 2 payment made. Most people pay either quarterly or annually, so would have between one and four payments. But it is possible to have many more.
- Class 3 – voluntary contributions. Like Class 2, there will be a record for each payment.
- National Insurance Credits – credits are allocated to people who are on contributory benefits (Incapacity Benefit (IB), Job Seekers Allowance (JSA) etc.). The credits are awarded to protect pension entitlement. Each spell on benefit is recorded separately.

These contributions data then form the base of the L2. For the purposes of easier analysis the L2 processing splits the different contribution types into separate datasets, so in L2 we have Class 1, Class 2, Class 3 and Credit datasets rather than a single contribution dataset.

## **Liabilities data**

In general the liabilities data holds all the information about when people become liable, or cease to be liable to pay National Insurance. The data is 'period' based, i.e. it holds start and end dates of particular activities or events. A number of different activities may run at any one time, and may overlap in terms of dates. In total there are over 75 different types of liability, some of which are no longer applicable, but are still relevant for past tax years, and some new liabilities that were not relevant in the past. The types of liability include:

- Abroad - Records periods spent out of the country, but only where the person notifies HMRC
- Arrivals - Records arrivals and re-arrivals from abroad (where the person informs HMRC). It also records the arrival and registration for National Insurance purposes of migrants from abroad.
- HRP – Home Responsibilities Protection is recorded on the NIRS account for the main payee of Child Benefit, and holds the dates of child benefit receipt. HRP protects the pension records of those undertaking caring responsibilities and is treated as a National Insurance credit.
- SEE – Small Earnings Exception. Periods of SEE are recorded indicating that the person is self employed, but exempt from paying National Insurance due to very low earnings/profits.

## **Methodology Statement for the Department for Work and Pensions L2 publications**

- RRA/RRW – Reduced Rate Authority/Reduced Rate Women. In the past married women could opt to pay a reduced rate National Insurance contribution and forfeit entitlement to benefit from their own pension. The liabilities data records the period of RRA/RRW.
- Class 2 & 3 – Records periods when Class 2 or 3 is liable to be paid.
- Class 1 abroad – Records periods when authority has been given to an employee to pay UK National Insurance whilst working in a foreign state, and whilst working for a UK employer.

## **Address data**

NIRS2 includes an address table, which records every notified change of address. Each address entry holds a start and end data. A new entry is created every time a new address is input to the system, which will include different spellings (for example, Road and Rd). Addresses on NIRS2 get updated from a number of sources, including the Customer Information System (CIS), DWP benefit systems (notified to NIRS via CIS), other HMRC systems and from customers themselves.

Experience with the NIRS2 data suggests that address data prior to 1997 is poor, and that post 1997 the data can only be regarded as accurate for people who have recently interacted with benefits, self employment, higher rate tax, or who have informed the HMRC directly. Having said that, the address data is improving significantly with the merging of HMRC's Computerisation Of PAYE (COP) system into NIRS2 and the integration of DWP's Customer Information System (CIS).

L2 includes the address data, although only the high level geographical coding, truncated postcode and start and end dates are currently made available to analysts for data security reasons. Currently L2 is not regarded as reliable below Local Authority level due to sample size.

## **Employer data**

NIRS2 holds two data tables holding the details of all employers operating a PAYE scheme. This information has only been recorded on NIRS2 since 1997. L2 includes 100% of this information .i.e. all employers operating PAYE since 1997 are included in the extract.

The EMPLOYER\_PRIMARY data hold all the details about the PAYE scheme, for example, the effective dates of the scheme, any changes to the employers PAYE number, and any changes to the employers address. Similar to the employee address data, a new entry is created for each change that the system identifies, for example Road and Rd.

Underneath the employer primary data sits the information about the PAYE return for each tax year, held in the EMPLOYER\_SCHEME table. This holds information about the amount of National Insurance, Tax and statutory payments paid to HMRC. It also includes information about the number of P14's covered by the PAYE return in each

## **Methodology Statement for the Department for Work and Pensions L2 publications**

tax year (giving a broad indication of employer size). In theory the employer data can be linked to all the employees at a P14 level. However, this is only possible for those cases that appear in the rest of the 1% L2 data.

In L2 the two employer data datasets are linked together. In addition, the audit trail is removed, the address information is run through postcode matching, and an anonymised dataset is made available to analysts.

## **Benefit data**

NIRS2 only holds benefit information relating to spells on contributory benefits where National Insurance Credits would be awarded to protect the pension entitlement record, for example, Jobseeker's Allowance, Incapacity Benefits and Carers Allowance. Information on the non-contributory benefits is not recorded, and thus does not make up part of the L2 extract.

However, L2 does include data on all benefits, which is added from the 5% quarterly extracts of the various DWP benefit systems. The relevant NINOs are extracted from the 5% extracts and are added to the L2 data. This enables the inclusion of start and end dates on all benefits back to 1997. This provides information on the following benefit spells: Income Support, Pension Credit, State Pension, Disability Living Allowance, Attendance Allowance and Widows Benefit.

## **Pension Scheme data**

The raw data from NIRS2 contains information on membership of second tier pension schemes and additional state pension.

Employees can build up an additional state pension on top of the basic State Pension; this was called State Earnings-Related Pension Scheme (SERPS) until April 2002 and is now called State Second Pension (S2P). S2P has more generous allowances for low and moderate earners, carers and people with illnesses or disabilities. Anybody earning less than the annual National Insurance lower earnings limit (£4,715 in 2009-2010), can still build up an entitlement if they:

- look after a child aged six or less, and are the person who claims and gets Child Benefit,
- take care of someone who is ill or disabled, or qualify for Home Responsibilities Protection
- are entitled to Carer's Allowance.

Occupational schemes are either contracted in schemes, where the member continues to pay National Insurance contributions in addition to any occupational pension contributions; or contracted-out schemes, where the member leaves the S2P by joining a contracted-out occupational, personal or stakeholder pension. Contracting out means you pay a lower rate of National Insurance but you lose or reduce your entitlement to S2P.

## Methodology Statement for the Department for Work and Pensions L2 publications

### Types of contracted-out schemes

- Contracted-out salary-related schemes (COSR) - Also known as a defined-benefit scheme gives members a final pension based on your salary when you leave scheme (some schemes use average salary), and your length of pensionable service
- Contracted-out money-purchase schemes (COMP) - Also known as a defined-contribution scheme gives members a final pension depends on the amount of money paid into the scheme, how well the money has been invested and how long you expect to be retired. Both employer and employee must make a minimum payment to pension, this is known as protected rights and equals the total of National Insurance rebates to employee and employer, HMRC also pay a yearly age-related rebate based on earnings between lower and upper earning limits. Most schemes need both employee and employer to make extra contributions on top of the minimum payment to boost the value of the fund. The employee can get tax relief on extra contributions.
- Contracted-out mixed-benefit schemes (COMB) - Employers can run schemes with COSR and COMP sections under the same scheme. The pension received depends upon which part of the scheme an individual is in.
- Appropriate Personal Pension (APP) – members pay full rate National Insurance contributions in the same way as S2P and the difference between this rate and the lower contracted-out rate (contracted-out rebate) is paid into a scheme which buys pension benefits at retirement age - protected rights.
- Stakeholder Pensions - available since 2001, members can pay contributions each year and basic rate tax (20%) is claimed on your behalf by Pension Company. This money is invested on the stock market, on retirement a quarter of capital accumulated can be taken out as a tax-free lump sum and the rest buys an annuity that pays the pension.

L2 holds four pension scheme tables that hold various levels of data on the information noted above:

- **Second Tier Pension Provision (STPP)** - Used to create the STPP tabulation tool. It is created from the Person, Scheme Memberships, Validclass1conts, SERPS members and S2P members datasets.
- **Scheme Memberships** - Contains details of occupational scheme memberships of individuals, including the start and end dates of scheme membership and scheme and employer contracted out numbers (SCON and ECON)
- **S2P members** - Contains information on who has built up entitlement to S2P and how they qualify for S2P, includes information on HRP (whether Child Benefit or Carers based) Earnings, and disability and illness related pension protection.
- **SERPS members** - Lists all SERPS members with their earnings in each tax year. Also has a SERPS indicator in each tax year and a SERPS duration (years). SERPS ended in 2001/02 tax year and was replaced by S2P in April 2002 so we expect these figures to remain constant.

**NIRS development timeline**

A brief timeline of the National Insurance recording System is detailed below

- 1948 NIRS system was introduced. All contributions held clerically
- 1974 Clerical records converted to computer. Some records not converted (already over State Pension Age or Pension Calculated)
- 1975 NIRS2 began. Summary information pre-75 was added.
- 1978 Qualifying period for pension purposes increased from 50 weeks to 52 weeks. SERPS and Contracted Out Occupational Pensions introduced
- 1987 NIRS amended to accommodate Personal Pensions
- 1988 NIRS amended to accommodate Defined Contribution pension schemes
- 1994 Development of NIRS2 begins
- 1996 Conversion of NIRS1 records starts. Some records archived at this stage
- 1997 NIRS2 Go live. Some processing backlogs and recording anomalies (which still exist). Some migrants registered without country of origin details.
- 1999 New procedures and rules for recording NI. Minimum wage & primary NI threshold introduced.
- 2004 New process for processing of PAYE Schemes introduced. Further backlogs in certain areas and subsequent data anomalies.
- 2005 Occupational & Personal Pension in payment P14 returns now included in NIRS2 in preparation for integration of NIRS2 and COP
- 2007 NIRS & COP (Computerisation of PAYE) amalgamated. Significant system changes to accommodate this change.
- 2008 Pension reform changes start to be implemented. Further changes to data.
- 2009 HMRC's National Insurance and PAYE Service (NPS) system went live in July 2009. Massive amount of cleansing needed by HMRC to close 're-opened' employments caused by platform change, after initial errors this should result in a more accurate system.