HM Government

Improving the completeness & accuracy of the electoral register through data matching

An evaluation of pilots testing data matching for the purposes of identifying new electors- *Full Evaluation*

July 2013

Cabinet Office
25 Great Smith Street
London SW1P 3BQ

Publication date: July 2013

Any enquiries regarding this document/publication should be sent to us at
electoralregistration@cabinet-office.gsi.gov.uk

Contents

Chapter 1
Introduction

1.1. Context

Under the current system of electoral registration an annual household canvass form is sent to each address, which is completed by one individual on behalf of everyone living at the property. From 2014 this system of registration will be replaced by one of Individual Electoral Registration (IER), with individuals making an application to register individually and providing personal identifiers (such as date of birth and National Insurance Number).

Ensuring that the registers are as complete and accurate as possible and that levels of completeness and accuracy do not decline under IER is a key aim of the Government. Data matching, whereby records on the electoral register are matched against other sources of public data, is one tool which could assist in ensuring that the registers remain as complete and accurate as possible, both during the transition to IER in 2014/15 and on an ongoing basis.

This paper presents the results of an evaluation of data matching specifically for the purposes of improving the completeness and accuracy of the register by identifying individuals who are not currently registered and inviting them to register. A separate evaluation, looking at the potential for data matching to confirm existing electors as part of the transition to IER, was published in April 2013[1].

1.2 Background

Completeness and accuracy of the electoral registers

The most recent estimates of completeness and accuracy indicated that following the annual canvass in December 2010 the electoral register was between 85 and 87% complete. This would mean that approximately 6.5 million people are missing from the electoral register. This compared to the best previously available estimate of

[1] https://www.gov.uk/government/publications/simplifying-the-transition-to-individual-electoral-registration

completeness of the registers in 2000 which suggested that around 3.9 million people or 8-9 per cent of eligible voters were not registered in 2000 (EC, 2005).

Completeness of the register has therefore declined over the past ten years, making it even more important that under-registration is tackled. Studies have also shown that certain groups such as young people (including attainers[2]), students, people who have recently moved house, people living in privately rented accommodation and/or shared households are less likely to be registered to vote.

The evidence suggests that the majority of inaccurate entries on the registers are related to people moving home and not informing the Electoral Registration Officer (EROs) (EC, 2011). It is worth noting that there is currently no requirement for people to notify the ERO when they move home, which makes it more difficult for them to identify home movers. Inaccuracies linked to fraud are thought to be relatively small in number (EC, 2010), and it has been suggested that levels of inaccuracy vary in line with levels of completeness.

What is completeness and accuracy?

The Electoral Commission (EC) defines completeness and accuracy of the registers as follows:

Completeness: 'every person who is entitled to have an entry in an electoral register is registered'.

Accuracy: 'there are no false entries on the electoral registers'.

Source: 'Great Britain's Electoral Registers 2011', EC, 2011.

Ensuring that the registers are as complete and accurate as possible and that levels of completeness and accuracy do not decline under IER is a key aim of the Government. In 2011 the CO ran an initial set of pilots exploring whether matching entries on the Electoral Register to other trusted public data sources could identify individuals who are not currently registered to vote but who may be eligible to do so, enabling EROs to contact these individuals and invite them to register.

These initial pilots took place across 22 local areas and tested data from eight separate data holding organisations. The pilots provided an opportunity to test the feasibility and processes for data matching. However, owing to the timing of the activities, which took place at the same time as the annual canvass, and the differing approaches of the pilot areas, it was not possible to distinguish whether new entries on the register were achieved as a result of the data matching or as a result of the usual canvass activities. It was therefore recommended that further testing was undertaken to enable a more robust analysis of the potential for data matching to improve the completeness and accuracy of the register.

The 2011 pilots were evaluated by the COand the Electoral Commission and the full evaluations can be accessed at:

https://www.gov.uk/government/publications/cabinet-office-evaluation-of-data-matching-pilots-2011

[2] 16/17 year olds who will become eligible to vote during the life of the electoral register.

http://www.electoralcommission.org.uk/__data/assets/pdf_file/0010/146836/Data-matching-pilot-evaluation.pdf

Data matching against national data sets

Building on the experience of the 2011 pilots and the evidence on levels of registration, the 2012 pilots were designed to focus specifically on three groups which are known to be less likely to be registered to vote and were identifiable within the available data sources, specifically:

- Recent home-movers
- Attainers (16/17 year olds who will become eligible to vote during the life of the electoral register)
- Students

The Statutory Instruments enabling the data schemes were The Electoral Registration Data Schemes Order 2012[3] and the Electoral Registration Data Schemes (No 2) Order 2012[4].

Two-tier data matching

Currently, to fulfil their registration duties EROs have access to a number of datasets to assist them in compiling the register and ensuring its accuracy. EROs are already legally entitled under Regulation 35 of the Representation of the People (England and Wales) Regulations 2001 and Regulation 35 of the Representation of the People (Scotland) Regulations 2001 to access any dataset held by the local authority that appointed them. They already make use of some or all of the following datasets[5]:

- the register of births and deaths;
- council tax records;
- registers of households in multiple occupation;
- local land and property gazetteers;
- housing benefit applications;
- lists of persons in residential and care homes; and - where allowed –
- details of “attainers” (those aged 16 or 17) held by education departments.

However, in England, only EROs appointed by Unitary Authorities are entitled to access to all of the council records for the area. In two-tier areas education records are not held by District Councils but by County Councils; EROs appointed by a District Council are therefore not able to access those records. A separate aim of the pilots was therefore to explore the potential value and practicability of opening the

[3] S.I. 2012 No. 1944

[4] S.I. 2012 No. 3232

[5] It is important to note that whilst EROs are legally entitled to access this data, not all will necessarily have agreed arrangements in place to do so.

legal gateway to enable data to be shared between upper and lower tier authorities in two-tier local authorities.

1.3 Research aims

The key aims of the COevaluation were to:

- assess the relative effectiveness of different data sets in identifying individuals who are not currently registered but may be eligible to do so;
- examine the number of new registrations that can be achieved by using the information obtained through data matching to invite individuals to register;
- examine the process of data matching and related implications for the effectiveness of any potential future roll-out of data matching for the purposes of finding new electors;
- explore the relative value of data matching as a tool for improving the completeness and accuracy of the electoral register; and
- examine the potential value of removing the legal barrier which currently prevents a lower tier authority (that holds the Electoral Register) from accessing the upper tier authority's data (that holds education records amongst others).

Chapter 2
Methodology

This chapter describes the methodology for the pilots, which was developed in close collaboration with researchers from the Electoral Commission which has a statutory duty to evaluate the pilots.

2.1 Participating areas

EROs from across England, Scotland and Wales were invited to participate in the 2012 data matching pilots. Areas were invited to partake in one or more of four pilot options, namely:

- Data matching targeted at finding recent **home-movers** using data held by the Department for Work and Pensions (DWP) and Royal Mail
- Data matching targeted at finding **attainers** (16/17 year olds who will become eligible to vote during the life of the register) using data held by DWP and education data
- Data matching targeted at finding **students** using education data and data held by the Student Loans Company
- **Two-tier data matching** exploring the potential for electoral services teams in lower-tier authorities to access data held by upper-tier authorities to improve the completeness and accuracy of the register

In total, 20 local areas volunteered to pilot data matching for the purposes of finding new electors, including 18 Local Authorities in England & Wales and two Scottish Valuation Joint Boards (VJBs). However, it is important to note that as these areas were self-selecting they cannot be assumed to be representative of all areas. The pilot areas and the options they selected to participate in are detailed in table 2a overleaf.

Table 2a: Overview of pilot areas

Local Authority/Scottish Joint Valuation Board[1]	Mining Option(s)	2011 mid year population estimates (16+)[2]
Barrow Borough Council (Cumbria County Council)	Two Tier	56,675
Ceredigion	Students, attainers and home movers	64,128
Conwy	Home movers	96,263
Coventry	Students and attainers	253,949
Greenwich	Students and home movers	199,927
Harrow	Attainers	192,324
Lothian JVB	Home movers	705,824
Mansfield District Council (Nottinghamshire County Council)	Two Tier	85,538
Pembrokeshire	Attainers and home movers	100,611
Powys	Attainers and home movers	110,310
Renfrewshire	Home movers	278,209
Richmond upon Thames	Home movers	150,419
Rushmoor Borough Council (Hampshire County Council)	Two Tier and attainers	75,385
South Ribble Borough Council (Lancashire County Council)	Two Tier and attainers	89,234
Southwark	Home movers	235,351
Sunderland	Attainers and home movers	227,315
Tower Hamlets	Students, attainers and home movers	205,645
Wigan	Home movers	258,205
Wolverhampton	Students and attainers	200,314
Wrexham	Home movers	109,228

Notes: 1) For authorities participating in the two-tier pilots the name of the upper-tier authority is provided in brackets.2) Source data Office for National Statistics, National Records of Scotland 2011 mid-year population estimates

2.2. Data sources used in the pilots

National Data Holding Organisations (DHOs) participating in Data Mining Pilots 2013

Department for Work and Pensions - Customer Information System (CIS)

DWP's CIS includes details of individuals appearing in databases kept by the Secretary of State for Work and Pensions for the purposes of social security. CIS is an amalgamated data source, consisting of information received from internal DWP heritage systems, as well as other government sources, such as Her Majesty's Revenue and Customs (HMRC). As a result CIS is seen within DWP as being the master of customer information.

The source CIS database is updated daily and includes a broad coverage of the population who are eligible to vote, including anyone who has been issued with a National Insurance Number (NINO). DWP Information Governance and Security Directorate extract data from CIS into its Warehouse to utilise for approved data matching initiatives. The CIS data extracted for the pilots included name, address and date of birth.

The data was used for both the attainers and the home-mover pilot options. For the attainers option data was targeted using date of birth so that only individuals who were between 16 years and 18 years of age (inclusive) were included. For the home-movers option only those individuals whose current address had been updated on the system within the last 12 months were included and a date of update was provided. Data was transferred from DWP to CO on encrypted disk using a secure courier.

Department for Education - National Pupils Database (NPD)

This data set is based on the data included in the NPD derived from the school census, which is completed termly in January, May and October. The data included details of individuals in maintained schools, academies and City Technology Colleges who were at least 16 years of age but less than 19 years of age at the date the information was included.
This data was used for both the student and attainers pilot options. Data was moved from DfE to CO using the secure 'Key to Success Platform'.

Welsh Education - Pupil Level Annual School Census (PLASC)
The PLASC collects data on pupil details, characteristics and curriculum and is updated annually[6]. Data included the names, dates of birth and postcodes of

[6] In addition data from the Lifelong Learning Wales Record (LLWR) was also provided. This data relates to learners across Wales submitted by learning providers via the LLWR. This data is used for funding, monitoring and analysis and provides the official source of statistics on Post 16 (non-higher education) learners in Wales and is updated daily. However issues with the format and quality of the data (e.g. levels of missing address data) prevented the data being used for matching and a decision was taken to exclude this data source from the pilots.

individuals but not their full addresses. Data collection for the PLASC takes place in January for all maintained schools and again in September for schools providing Post-16 education.

This data was used for matching in both the student and attainers pilot options: however as the data did not include full addresses but only postcodes, it could not be used by local authorities and therefore no results for this data set have been included in this evaluation report[7]. The secure Data Exchange Wales initiative (DEWI) was used to transfer files to Cabinet Office.

Student Loans Company (SLC)

The SLC data was extracted from the Student Finance Customer Account System (Higher Education) and included current students' name, address, date of birth, address where application was made and term-time address. The copy database is updated weekly..

This data was used for the student pilot option and was transferred to CO by secure email.

Royal Mail

Data corresponding to the 14 local areas seeking home movers was selected from the National Change of Address (NCOA) Update database. This includes the names and both new and old addresses of individuals who have redirected their mail and given their permission to pass on their information to selected companies and other organisations who already know them to update their contact details. Also included was data from the NCOA Suppress database which is from the Royal Mail's Redirection Service and includes the name of the individual who has moved and the address they have moved from. This data is used to stop companies mailing to people who have gone away. Both data sets receive additional records on a daily basis.

Results derived from the matching process were sent to the via secure disk and courier.

Two-tier data sources

Whilst a range of available data sources were discussed, based on the available data sets a decision was taken to focus on data sets that specifically targeted attainers for the purpose of the pilot. The range of upper-tier authority data sources employed in the two-tier pilots included[8];

- Name, address and date of birth of individuals appearing in databases kept by the Council relating to education, who are pupils at schools maintained by the Council and who are at least 16 years old but less than 19 years old;

[7] When the pilots were originally planned it was hoped that this data might be used in conjunctions with other sources – locally or nationally – to help identify people, or that areas might use it to work with schools. However, due to time constraints this was not possible.

[8] These are the data sources that the EROs were legally permitted to access for the pilots although not all areas used the full range of data in practice.

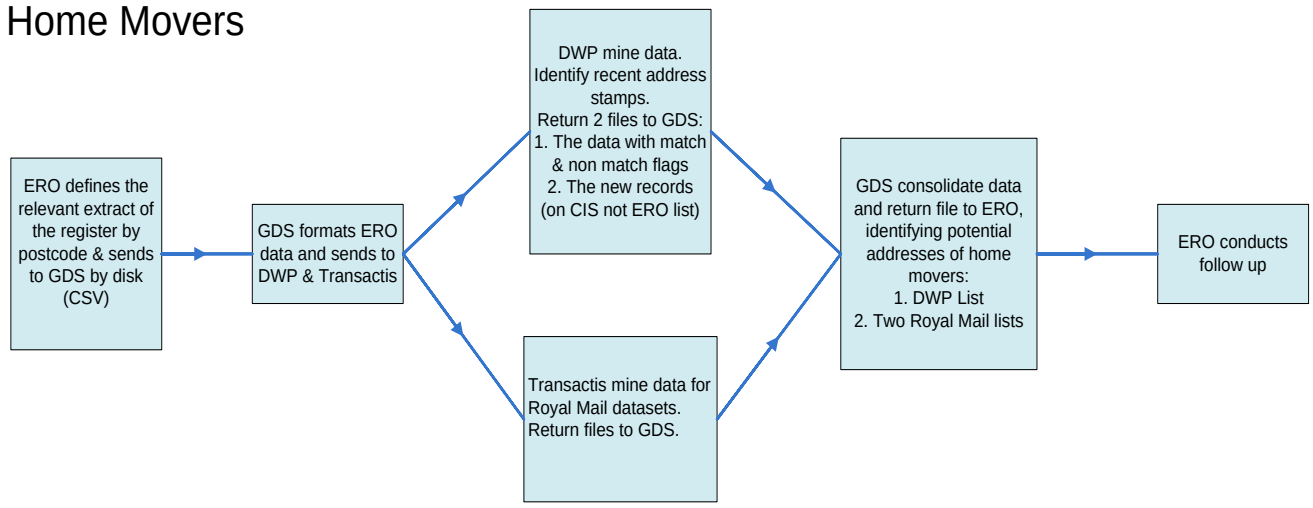
- Name, address and date of birth of individuals with special educational needs or a disability, placed by the Council in schools recorded in the Council's education database and who are at least 16 years old but less than 19 years old;
- Name, address and date of birth of individuals who are recorded in the Council's children's social care database as looked after by the Council and who are at least 16 years old but less than 19 years old; and
- Name, address and date of birth of individuals appearing in databases kept by the Council for the purpose of recording participation in education, training and employment and who are at least 16 years old but less than 19 years old.

2.3. Data matching process

The primary legislation[9] which gives power to set up data schemes requires them to be based on proposals made by local registration officers. However, based on the learning from the 2011 pilots, participants were requested to follow the same broad process. Figures 2a-c provide an outline of the broad process adopted for the pilots using national data sets.

As the aim of the two-tier pilots was to test the feasibility of data matching and explore how the process might work, we did not set out a specific approach for matters such as the matching and transfer of data. Instead, we worked with the individual pilots to agree processes between the upper and lower tier authorities, which are discussed in the findings of the report.

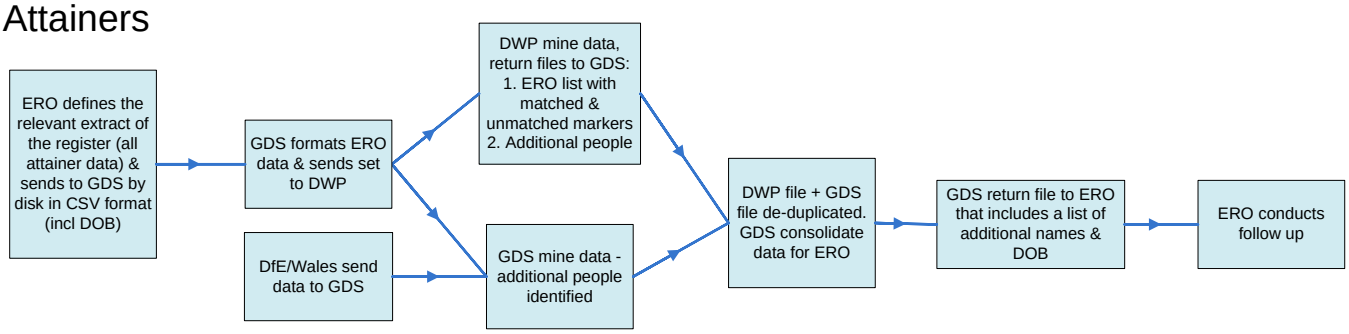
Fig 2a-c: Process maps for pilots using national data sources



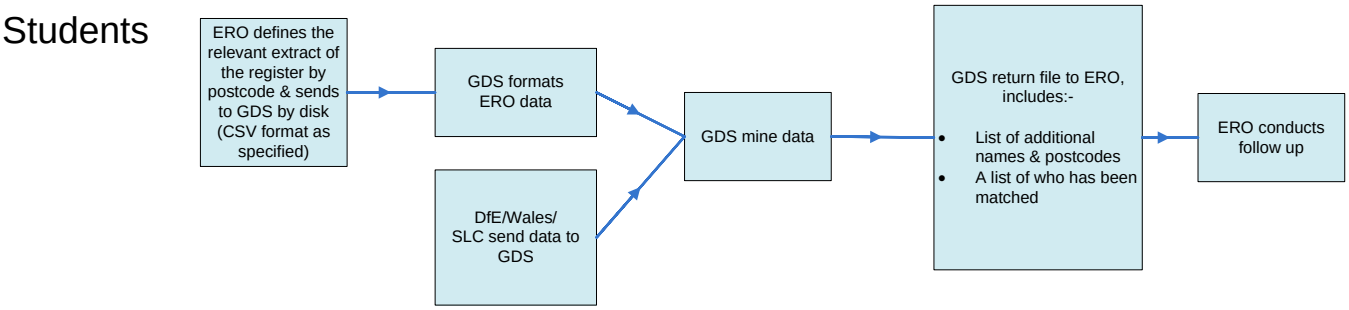
Notes: 'Transactis' are the data processor for Royal Mail

[9] The Political Parties and Elections Act 2009, section 36(1).

Attainers



Students



Matching process

In order to comply with legal restrictions in relation to their data, Royal Mail/Transactis completed the matching process internally using their existing matching algorithms, ensuring that data was only supplied to the CO orEROs where they had appropriate permission to do so. Home-mover data was matched at DWP using the pilot matching algorithm from the 'Confirmation' data matching pilots[10]. The CO Government Digital Service (GDS) undertook the matching for all data for attainers and students using an algorithm that was designed to mirror the matching principles used by DWP for the home-mover data. It is therefore important to note that while the matching process will have followed the same principles, there were some differences between the exact matching algorithms used.

Return files to EROs

The files returned to EROs included only those individuals which the matching process had identified as not currently on their electoral register, with the following record level details:

- Name
- Date of birth (available for all data except Royal Mail)
- Full address (as it appeared in the original DHO data)

[10] Full details can be found in Annex A of the full evaluation of the Confirmation Pilots https://www.gov.uk/government/publications/simplifying-the-transition-to-individual-electoral-registration

- Currency marker (only available for DWP home-movers data, this marker indicated whether DWP had a record of an address update within 0-3 months, 3-6 months, 6-9 months or 9-12 months).
- Number of data sources on which records were located and a 'true/false' marker for each data source, enabling the ERO to ascertain in which DHO's data the individual's details had been found.

ERO follow up

Local data checks

Whilst every effort was made only to return to EROs records of individuals who were not currently registered, it was known that due to the limitations of automated matching some records of people who were registered may have been included in the data. In addition it was known that the data would include some individuals who are ineligible to vote (for example because of their nationality). Therefore guidance provided to participants in the pilot recommended that all areas conduct additional checks of the data provided against their local register. Participants were also invited to check their data against other locally held data sets where they had the capacity and capability to do so

Invitations to register

However, to ensure consistency for the purposes of evaluation all areas were asked to follow the same broad process for following up potential electors identified in the data matching. This included conducting at least one write-out to potential new electors identified during the data matching, and a template invitation letter was provided to EROs. Areas could then choose to conduct any additional reminder write-outs or canvassing dependent on their capacity.

2.4. Evaluating the data matching

This report is based on feedback and data obtained from a range of sources including:

- CO led workshops which provided an opportunity for the CO to: update pilot areas and other relevant parties (e.g. DHOs) on progress; to gain feedback from attendees; and to provide a forum for pilot areas to share their experiences with other participating areas.
- Qualitative interviews conducted over the telephone with each of the pilot areas towards the end of the pilot to examine in more depth the views and experiences of pilot areas and to gain further insight into lessons that could be learnt for the future. These interviews were recorded for accuracy and analysed using a thematic matrix.
- Standardised reporting forms submitted by pilot areas at the end of the pilot which provided information on the number of records checked on local data,

the number followed up via write-out and personal canvassing and the responses received.

- Summary data from GDS relating to results of the matching undertaken by GDS.

Chapter 3
The process of data matching

This section describes the overall findings from each stage of the pilot, including key contextual information that should be considered when interpreting these findings. The findings from the data matching involving national data sets and the two-tier data matching pilots are provided separately.

3.1 Process issues identified in matching against national data sets

Process and logistics prior to data matching

Before any data sharing could take place between the DHO and the local areas (LAs) there were a number of standard requirements, essential activities and documents which all pilots were asked to complete or have in place due to legal requirements (i.e. Statutory Instrument or the Political Parties and Elections Act 2009). These included:

Article 4 Agreements (Information/Data Sharing Protocols)

As in the statutory instrument that enabled the 2011 data schemes[11], Article 4 of each of the Electoral Registration Data Schemes Order 2012 and the Electoral Registration Data Schemes (No 2) Order 2012 required every participating ERO to make a written agreement with every DHO with which they were to match data. These agreements contained detailed information as to the respective obligations of the ERO and the DHO and set out the exact basis for the processing of data, including the requirements for the transfer, storage, destruction and security of data and the consequences of failing to meet those requirements.

The agreements were required to be signed by the ERO for each pilot site and by an appropriate official of the DHO concerned (normally at Director level). As a number of data holding organisations had consented to the CO matching their data on their behalf, the agreements concerned were also required to be signed by the Programme Director for the Electoral Registration Transformation Programme at the CO.

[11] The Electoral Registration Data Schemes Order 2011 (S.I. 2011 No. 1466)

Completed Privacy Impact Assessment

Each of the pilot schemes was also subject to its own privacy impact assessment setting out the details of the scheme, its effects upon individuals, potential privacy risks, security measures and compliance with the Data Protection Act.

To facilitate this process the CO centrally co-ordinated the development and distribution of Article 4 agreements in order to agreed their terms and obtain the relevant signatures. This meant that all pilot areas were provided with Article 4 agreements for signature which had already been agreed with and signed by the DHOs with whom they were matching their registers. The CO also supplied template Privacy Impact Assessments for the schemes for final completion by the pilot areas. Whilst the pilot areas themselves reported finding this process relatively simple to complete, it placed a significant administrative burden on the CO team. This was particularly the case for Royal Mail data where the data sharing requirements specified the use of wet signatures on a single document as well as the completion of additional 'end user' agreements that participating authorities were required to sign[12].

As a result of the complexities of getting these legal agreements in place, the overall timescales for the matching and follow up work were reduced. This had a subsequent impact on the remaining pilot activities. For example, it limited the time that GDS had to test the data matching and output files, and many of the pilot areas which had planned to carryout personal canvassing in addition to the write-out were unable to do so.

If future data matching exercises were to take place on a wider scale sufficient resource will need be centrally allocated to ensure that the appropriate documentation is in place to meet the legal requirements for data sharing, although steps may be taken to reduce this burden. For example, the practical challenges of obtaining multiple signatures on the Article 4 agreements was similarly identified in the 2011 pilots and as a result the 2013 legislation, which has already enabled data matching for the purposes of Confirmation[13], was drafted to enable equivalent data sharing instructions to be put in place without requiring signatures (unless, for example, a signed agreement is explicitly required by the DHO). This significantly reduced the resources required for co-ordinating these agreements. (Insert footnote - As the current data matching pilots were enabled under the same legislation as the 2011 pilots it was not possible to implement any changes to the process).

Processing and matching the data and producing the match reports

As described in the previous chapter whilst the process for matching the different data sources differed between data holding organisations (in order to meet the legal

[12] Including an end user letter on headed notepaper signed by the ERO.

[13] Confirmation' is a data matching exercise that is being undertaken as part of the transition to IER whereby existing electors' names and addresses on the electoral registers are matched against records held by DWP in order to verify the identity of people currently on the registers. Where an existing elector's details can be confirmed by the DWP, the ERO can transfer that person to the IER register.

requirements around releasing data) all data was centrally processed and collated into files by GDS prior to the data files being sent to the EROs.

Based on the findings from the 2011 pilots the format of the data returned to the EROs was simplified so that only the details of the individuals who had been identified through the matching process as not currently on the electoral register were returned to EROs. In addition, as part of the processing the files were de-duplicated meaning that an individual was only returned once within the data set even if they had been found on more than one data set.

Overall, this process appears to have worked well, and areas that participated in both the 2011 and 2012 pilots reported significant improvements in the format and accessibility of the files returned to them. Furthermore the vast majority of the areas reported that they were happy with the format and content of the files. However, feedback from both GDS and the pilot areas highlighted a number of outstanding issues with the data matching process and ways in which this process could potentially be further improved. These are discussed below.

Specifications of file formats

Whilst the requirements for each DHO detailed the general format and content of the data fields required, GDS reported that **agreeing more detailed technical specifications for the format of the data with DHOs at the outset of the pilot may have been beneficial in reducing the resource required to process the data.**

Increased time for testing and reviewing files

Owing to the delays in completion of the legal agreements required to enable the data to be shared, GDS had a limited window in which to process the data if it was to be transferred to the participating EROs within the legislative time frames[14]. Whilst GDS had built and tested the process using dummy data, this limited the time that they had to test and refine the matching using live data. In addition this prevented any user testing of the files (whereby initial files are sent to a small number of areas for review) which had been found to be beneficial in the earlier 'Confirmation' data matching pilots, not only for refining the matching and file formats but also for informing the supporting guidance provided to EROs.

Therefore, **if similar data matching exercises are to take place in the future sufficient time should be built in to the process for processing and testing the files.**

Data quality issues

The following section describes some of the issues with regard to the quality of the data for the purposes of data matching in order to find potential new electors. **It is important to note that these issues are not necessarily reflective of the quality of the source data,** as issues identified are not necessarily problematic in relation to

[14] The statutory framework that enabled the data sharing included a requirement for all data transfers to have been completed by a specific date.

the original purposes for which the data is collected (e.g. inclusion of individuals who are ineligible to vote for nationality reasons but who are quite properly present in the source data). In addition, some of the quality issues discussed arose from the matching process itself, as opposed to the source data.

Duplicate entries within the data

As described previously, as part of the processing of the data GDS attempted to de-duplicate between data sources for each file. However, as the files for each option were processed separately, where pilots opted to conduct more than one option the results were not de-duplicated across options. This resulted in some areas receiving duplicate results across files. In addition, whilst the GDS processing aimed to identify and remove entries that appeared on more than one data source in an option the process was not designed to identify duplicate entries within the DHO data sets and feedback from the pilot areas suggested that this was an issue with the data. This indicates that **adding additional steps to the processing of the data to further identify and remove duplicate entries both within and across options would be beneficial for any future exercise**.

Quality of address information

Reports both from the individuals processing the data at GDS and from pilot areas receiving the data highlighted that, with the exception of Royal Mail data, the address quality of the data presented to them was poor. Common issues reported included missing flat numbers, the use of historical county names and the use of abbreviations in the data[15]. This had implications in terms of the time taken to process that data and the effectiveness of the matching in the initial matching stages, as well as increasing the time taken to process the results at the ERO level (both in terms of checking records against the electoral register or other locally held data sets, and also in relation to amending some addresses prior to write out).

As the electoral register is a property based database and is regularly used for the purposes of writing out to individuals, the quality of address information is relatively high. By contrast, not all DHOs use their data to write out to individuals and therefore the overall quality of address information has less significance for the purposes of the data held. For example, for individuals contacting DWP or one of the other organisations feeding the DWP CIS database, address information will be updated manually based on the address details given during contact. Where there is no subsequent need to write to the individual, errors in the address information may not be picked up.

The poor quality of address information was also highlighted in the 2011 pilots, and as a result of the learning from that exercise DWP has added Unique Property Reference Numbers (UPRNs) to their data[16]. The availability of these UPRNs facilitates the matching process and the ease with which areas can check the data and was positively received by pilot areas as it provides a consistent reference

[15] Other errors reported included inclusion of overseas addresses and commercial property addresses.

[16] UPRNs are standardised unique identifiers for each land and property unit and are heavily used by EROs to conduct their current activity. UPRNs are assigned to every unit of land and property recorded by local government who have a statutory obligation to record all changes in property details, including all new builds.

between the addresses within DWP CIS and the electoral register. However, whilst the majority of DWP CIS records have been populated with UPRNs, not all records have them; and the addition of the UPRN has not resulted in amendments to the textual address fields which were returned to EROs as part of the data[17]. Furthermore, other DHOs do not have UPRNs on their data as it may not relevant for the purposes for which they use the data.

The addresses returned to EROs were taken directly from the DHO source data. If a similar exercise were to be carried out in the future **one potential option to resolve address quality issues may be to add an additional step to the matching process whereby the address information is matched to the property list held by each ERO, enabling the address information provided by the ERO to be returned to the ERO as part of the data**. However, it is important to note that the data is made up of records that could not be matched to ERO data and in some cases this mismatch will have arisen because they failed to be matched at the address level.

Information on the proportion of records that mismatched on address is not currently available and therefore it is not possible to predict the proportion of records to which this would apply. Consideration would need to be given as to whether to exclude these records from the data returned to EROs or to provide them with the source DHO address only. **Limiting the return data to only include records where the address can be matched to the ERO property list but where there is no identity match could be one way to provide a more targeted data set, although it may result in the exclusion of some genuine potential new electors[18].**

Inclusion of individuals who are ineligible to vote

Not all individuals included within the source data sets are eligible to register to vote, for example owing to their age or nationality[19]. Whilst in the majority of the DHO data sets availability of date of birth (DOB) made it possible to identify the age of individuals, not all data contained DOBs. This was notably true of Royal Mail data and resulted in some letters being sent to individuals below registration age during the pilots[20].

Information on nationality is not available on any of the DHO source databases. As part of the manual checking of records against the electoral register a number of EROs were able to identify where they had a record of an individual being ineligible on the basis of nationality in order to exclude them from any write-out, but EROs will only have a record of this where individuals have previously informed them of it.

---

[17] Overall DWP estimate being able to allocate UPRNs to around 95% of records, however it is not possible to successfully match all addresses to the centrally held data on UPRNs and differences in the frequency within which UPRN data in DWP and the local registers are updated will also lead to some discrepancies between data sets. It should also be noted that whilst the majority of electoral services teams store UPRNs on their data not all necessarily will.

[18] An identity match is a match of the name details of the record.

[19] Further information on eligibility to registered to vote can be found at http://www.aboutmyvote.co.uk/who_can_register_to_vote.aspx

[20] Royal Mail have subsequently informed the CO that they are confident they would be able to exclude individuals below registration age.

This has two broad implications for any future use of data matching to identify potential new electors. **First it emphasises the need for EROs to conduct additional checks on the data returned to them prior to writing out to individuals. Second it highlights that it will not be possible completely to prevent invitations to register being sent to individuals who are not eligible to register, which may cause confusion for the citizen**. The risk of confusion for the citizen might be mitigated by ensuring that the letter or information sent to the individual provides a clear explanation of why he or she is being contacted and the fact that receiving the letter does not mean that the recipient is necessarily eligible to vote. Information on eligibility should also be included.

Data collected in the qualitative interviews showed that pilot areas had mixed views on the impact of inclusion of ineligible individuals in terms of the usability of the data. Some areas reported that they would not be comfortable using the data to contact individuals unless they could be confident that the details of individuals who are ineligible to register were excluded from the data. Other areas reported that there may still be some benefit in contacting these individuals, as it enables them to update their own data to identify households where the residents are ineligible.

'Inaccuracies' identified in the data

Pilots reported that, based on a combination of their manual checking of the data against their local data and responses from the public to the follow-up activities, a number of inaccuracies were identified in the data. These inaccuracies occur where an individual is recorded on the source data as resident at a particular address but the ERO has other intelligence showing that the person no longer resides in the property, or as a result of the pilot activity the ERO has received information from a property to the effect that the person contacted is not resident at the address.

One of the concerns highlighted by participating authorities is that if data matching is to be used on an ongoing basis **a mechanism would be required for recording and/or providing feedback to the DHOs regarding the inaccuracies identified**. This would be necessary in order to prevent EROs from having to re-check the same data or (where an individual has contacted an ERO to report an inaccuracy) contacting citizens with the same inaccurate information on multiple occasions.

Consideration will need to be given to the most appropriate way to manage this. One option would be to report inaccuracies to DHOs, enabling them to update their own data. However this will have resource implications for DHOs; and it may not be feasible for the DHO to alter individual details without written confirmation from the individual, which may have legal and practical implications(for example where an ERO is informed that someone has left an address but no forwarding address is available). An alternative option would be for this information to be stored within the ERO's electoral management software (EMS). These **options will require feasibility testing including consideration of legal issues around storing such data, particularly where information has been obtained through a third party**.

The existence of these inaccuracies further reinforces the principle that information from the DHOs should only be considered as one form of intelligence. It will not possible to guarantee that all the DHO information is correct and therefore this data should be viewed alongside other intelligence.

Other data issues

In addition to the issues identified above other limitations of the data were identified, specifically that:

- Royal Mail data will inevitably include records where an individual has selected an address as a forwarding address although they may not be living there.
- A number of deceased individuals were included within the data returned to EROs. Even in the case of DWP CIS data, where the data sets contain a record of an individual being deceased, it is acknowledged that DWP may not become aware of a death in all cases and the ERO is likely to have access to the most up to date information on registered deaths through the local registrar.
- One local authority identified the inclusion of some addresses that were deemed 'sensitive' (e.g. Women's Refuges and rehabilitation hostels). Whilst some sensitive addresses are suppressed in the source data, feedback from DHOs suggests that it will not be possible to guarantee that these addresses are not included within the data.

It will therefore be important to ensure that EROs are provided with clear guidance on the data sets that enables them to understand the relative limitations of the data sets, and the recommended checks of the data that may help address some of these limitations. The extent of any issues arising from these limitations would also need to be closely monitored.

Limitations of automated matching algorithms

As part of the process of the pilots, and based on feedback from the 2011 pilots, the aim was to only return to EROs records of individuals whom the data matching process indicated were not on their register. However, it is recognised that in some cases records may mismatch due to differences in the format or spelling on the different data sources, as opposed to being genuine mismatches. Whilst further refinements to the matching algorithms used in the pilots will be possible and may reduce the number of inaccurate mismatches the implication of this is that **some manual checking will still be required to prevent EROs contacting individuals who are already registered to vote[21].**

Responses from the public

As part of the qualitative interviews pilot areas were asked whether they had received any enquiries or feedback about the pilots from members of the public. Participants reported varying levels of queries or contacts from the public. Some areas reported

[21] The matching algorithm used by DWP was the 'pilot algorithm' used in the Confirmation data matching pilots which has subsequently been refined to improve the matching capability. Full details are contained in Annex A of our evaluation of the Confirmation pilots https://www.gov.uk/government/publications/simplifying-the-transition-to-individual-electoral-registration

that they had received few or no queries, whilst others reported a short peak of contacts a few days after the initial write-out was sent. As would be expected, areas that sent out larger volumes of letters were more likely to report larger volumes of contacts indicating that **if data matching exercises were to be conducted on a larger scale EROs would need to be prepared to manage enquiries from the public which are likely to peak just after the write-out**.

Participants reported that a common cause of contact was where an individual had received a letter at their property addressed to someone who did not live there, with some reporting that the individual being contacted had never lived there or had moved away some time ago. In some cases people were concerned about why the ERO had incorrect information and where it had come from, and what they could do to inform the source DHO that the information was incorrect. As described earlier, it is currently unclear if or how it would be possible for DHOs to amend or update data based on third-party information. However, if similar data matching exercises were to be conducted in the future **EROs are likely to require guidance on what they can advise individuals in these circumstances.**

Identifying potential fraud

By providing EROs with additional sources of intelligence on the residents of a property, data matching can provide an opportunity to improve the accuracy of the register by identifying individuals who are currently registered at an address but do not reside there, as well as by identifying individuals who are not currently registered. In most cases individuals being registered but not resident at a property will be a result of residents moving out and EROs not being aware of this, rather than an indicator of fraudulent registrations. However, it is possible that the data could help to identify the small minority of cases of fraudulent registrations.

Whilst the 2012 pilots did not set out to test this specifically, as part of the qualitative interviews pilot areas were asked about the potential for the data to be used in this way. **Participants reported that any data that helped to build a more complete picture of the residents of a property is useful and can be used as an indicator of potential inaccuracies or fraud to be considered alongside other information**. **However, most areas also reported that they did not perceive electoral fraud to be an issue in their area and therefore felt that the data would have limited use specifically in relation to identifying electoral fraud.**

In addition to providing a mechanism for identifying possible electoral fraud this type of data matching exercise also has the potential to indicate other types of fraud. For example, a number of EROs noted that they had identified information indicating that individuals may be inaccurately claiming a single-person discount for Council Tax. **The primary purpose of the data matching is to improve the completeness and accuracy of the register: however the pilots have demonstrated that there may be cases where the exercise provides evidence of non-electoral fraud and therefore clear guidance would need to be provided to EROs and DHOs on what action they are permitted and/or required to take in these cases.**

Resource implications for electoral administrators – a manual vs. automated process

Overall pilot areas reported finding the broad process of the pilot exercise straightforward. The vast majority of areas reported that the format and content of the data files returned to them were simple to interpret and analyse[22]. As described above, pilot areas reported that the most resource intensive part of the process was the checking of the data against their local register and/or other locally held data sets, which for the vast majority was completed manually. **The majority of areas fed back that the process would need to be more automated in order to reduce the burden on their time** of manual checking including by:

- Enabling checks against other local data sources to be automatically processed
- Enabling data to be integrated into their existing EMS systems to facilitate cross-referencing between data sets

However, as the results described earlier highlight, even where data has been passed through an automatic matching process (like the data returned to EROs in the pilot) there will remain a proportion of records that, due to differences in spelling or format of data between data sources, could not be matched by an automated matching programme. Therefore, **some manual checking of records will always be required to avoid sending invitations to individuals who are already registered.**

In addition to supporting EROs in conducting checks of the data returned to them, **there is also the potential for greater automation of the process, through integration with their existing EMS systems, to be beneficial in terms of the resources required to followup records**. For example, a number of areas highlighted that sending out invitations to register through their EMS would enable them to send out bar coded invitations to register making it simpler for them to input the details of any registrations achieved as result of the data matching.

Resource implications centrally

In addition to the resource requirements for EROs, data matching in this way also presents resource implications centrally. The experience of the pilots illustrated that if data matching for the purposes of finding new electors were to be rolled out, resource would need to be allocated centrally both to manage the initial set up and on an on-going basis.

The actual requirements will vary depending on the scope of any roll-out. **In addition to any initial costs of setting up the process, some of the key potential ongoing resource requirements identified in the pilots include:**

- **Co-ordination with DHOs and funding relevant costs to the DHOs**. Costs may vary according to DHO. For example, in the pilots only Royal Mail charged the CO for the use of their data as all other DHOs were other government departments. Other costs to DHOs include legal and related

[22] Some issues were reported with the CSV text format, however these were resolved relatively easily when areas received guidance internally or from the CO on how to convert this format into Microsoft Excel.

costs to ensure that all data is shared within the limits of the legal requirements enabling the data sharing. This co-ordination will also include reviewing any changes to the source data and ensuring relevant agreements are in place to enable data sharing.

- **Co-ordination with EROs**. This will involve ensuring that EROs are provided with clear guidance and, if data matching was offered on a flexible basis, administering applications for the data and acting as a central point for any queries. It would also involve ensuring that any guidance produced is maintained and updated.
- **Facilitation of secure transfer of data**. In the pilots, data was securely transferred in a number of ways such as secure online platforms and secure email. In some cases it was necessary to transfer the data through secure courier and encrypted disc, as not all EROs had access to a secure email[23]. Ensuring a consistent and effective way to transfer the data will be important and is likely to require administrative resource.
- **Matching the data and processing the files.** Once the processes for matching have been set up and tested there will remain an ongoing requirement for the matching against each DHO to be completed and for the results to be collated into a single return for each ERO. In the pilot the approach to matching varied, with some DHOs conducting the matching themselves whilst others supplied their full data which was then matched by the CO (GDS), however consideration would need to be given to the most appropriate approach for the future.

Participants' views on the future of data matching

Views on the value of data matching

As part of the qualitative interviews, participating authorities were asked a series of questions around the potential future for data matching. **Overall, the majority of the pilot areas report that having access to the data was valuable and that they would wish to use the data again given the opportunity**. However, many areas reported that for the process to be practically viable many of the data quality and process issues discussed previously would need to be addressed. It should also be noted that as the pilot areas were self-selecting they are likely to have a greater interest in the matching and therefore cannot be considered as representative of all authorities.

Participants were also asked about the perceived benefits of data matching using national data sets over other types of targeted registration. **A key issue arising from these responses was in relation to their ability to access similar information through the use of local data sets, and a number of areas reported**

[23] Transferring data through secure courier is relatively costly, however as part of the preparation for the introduction of IER all EROs will be required to have access to a secure email account and/or secure access to the IER digital service (which provides a potential mechanism for transferring data securely). The potential to use these alternative means of securely transferring data may negate or at least significantly reduce any need to use secure couriers in the future but would require full feasibility testing.

that the exercise had encouraged them to be more proactive in using the data they already had available. As discussed earlier the results of the pilot indicate that a significant proportion of individuals identified by matching against national data sets could also be located in local data sets. However, the national data does still appear to add additional value in terms of identifying additional potential electors and may also be a useful mechanism for providing the ERO with additional confidence in the accuracy of the local data where an individual appears on both data sets.

As previously discussed, participants were also concerned about the scalability of the process given that they found the manual process of checking the results against their register and other local data extremely resource intensive. A number of areas suggested that this type of activity would only be feasible if it remained targeted on particular groups or areas to ensure that the volume of records remained manageable. **This suggests that if this data were to be made accessible for this purpose in the future this should be voluntary, with EROs opting-in based on their own assessment of the relative value this type of data matching would have in their area and the groups or areas that they wish to target**.

Views on the timing of data matching

Some pilot areas also highlighted that if data matching is to be used it may be **particularly beneficial prior to the confirmation process, as individuals who provide their details in response to the invitation ahead of the confirmation exercise would effectively already have been confirmed through the data**[24]. In addition there is the potential to combine the work with other local matching activities where these are being planned as part of the confirmation exercise.
If data matching for the purposes of finding new electors were to be introduced **as business as usual, the preferred timing for the activity is likely to vary according to the groups targeted**. Pilot areas reported that students may be most usefully targeted at the start of the academic year in September or October. In the case of home-movers and attainers, participants consistently reported that using data matching as part of the annual canvass process would be most effective: however, views differed on whether the process would be most usefully targeted either just prior to the annual canvass or just after it. This suggests that **a flexible approach to the timing of the data matching would be beneficial, although this would need to be considered against any related resource implications in terms of the management of the process.**

[24] 'Confirmation' is a data matching exercise that is being undertaken as part of the transition to IER whereby existing electors' names and addresses on the electoral registers are matched against records held by DWP in order to verify the identity of people currently on the registers. Where an existing elector's details can be confirmed by the DWP, the ERO can transfer that person to the IER register. In addition, EROs may conduct matching against additional local data sets to match and 'confirm' existing electors who could not be matched within the DWP data set. It should be noted that this benefit would not apply post transition to IER as the data sets used are unlikely to meet the requirements for verification matching.

3.2 Two-tier data matching

EROs have a statutory “right to inspect records kept in whatever form by the council by which [they have been] appointed”. In unitary authorities EROs are legally entitled to access relevant council datasets (as detailed on page 5), however, in a two-tier local government area a legal barrier means a lower tier authority holding the electoral register is unable to access the upper tier authority’s data (which will include education records, amongst others).

As part of the current pilots, the CO worked with four lower-tier authorities to test the usefulness and practicability of allowing local authorities in two-tier areas to do internal data matching by removing this legal barrier. The aim of this piloting was to test the process and practicality of doing so.

Connecting with the UpperTier

The strength of the relationship between the upper and lower tier authorities involved in the pilots varied: some had worked with (and were co-located with) the upper tier, whilst others had not had any previous contacts. A number of areas reported that the involvement of the CO had assisted in ‘bringing [the upper tier] to the table’.

However, whilst pilot areas reported some initial challenges in establishing the right contact within the Upper Tier authorities, once the relationships had been established they appeared to work well. All areas reported that they found their contacts in the upper tier authorities helpful in identifying and establishing the potential data sources that they could use.

Transferring data and conducting the matching

Once the data sets had been identified some areas experienced challenges in identifying a secure method of data transfer although in all cases a solution was identified (for example secure email). Once the data was received, pilot areas reported similar issues with the quality and limitations of the data as were experienced with the national data sets. However, a key benefit of using two-tier data (or other local data) is that ERO has direct contact with the data holder which can facilitate their understanding of the strengths and limitations of various data sets.

In all four areas the actual matching was undertaken by the lower-tier authorities and completed manually (Barrow used a combination of manual and automatic matching). Participants reported that this was resource intensive and that being able to conduct the matching on an automated basis would be beneficial. Whilst the aim of these pilots was to explore the process and feasibility of the matching, some areas did opt to conduct follow-up work with electors, where comparable data is available this is presented in the following chapter.

Views on the benefits of two-tier data matching

Overall participants were positive about the use of the data and its potential for identifying new electors, although some participants raised concerns about the

scalability in relation to the potential for upper-tier authorities receiving multiple requests from lower –tier authorities, suggesting that some level of co-ordination between authorities may be beneficial. .

It is important to note however that, as is the case currently with unitary authorities, having a legal entitlement to access data sets in and of itself will not guarantee that all areas will put this legal entitlement to use.

Chapter 4
The potential impact of data matching on the electoral register

This section describes the results of the data matching pilots including: the volume of potential electors identified in the DHO data; the results of subsequent checks of the data by pilot authorities; and the response rates to the invitations to register sent out by the participating authorities.

As the two-tier element of the pilot was primarily focussed on exploring the process and the feasibility of removing the legal barrier that currently prevents lower tiers accessing data that is available to EROs within unitary authorities, the results presented in this section primarily focus on national data sets. However, relevant data from the two-tier pilots is presented where available.

The complete data sets upon which the findings in this chapter are based are available as a separate data tables alongside this report on gov.uk.

4.1 Overall match results

Home-movers

The EC estimates that the register declines in accuracy by, on average, approximately ten per cent over the course of a year, primarily due to people moving home. This will vary between areas depending on the relative mobility of the population they cover. Royal Mail data will cover only a relatively small proportion of these movers who opt to pay for a redirection service and for whom Royal Mail have the appropriate permissions to share their data. Across the 14 areas participating in the home-movers pilot, the volume of records returned by Royal Mail equated to between 0.6 per cent of the total register size in South Ribble to three per cent in Renfrewshire (average 1.5 per cent)[25].

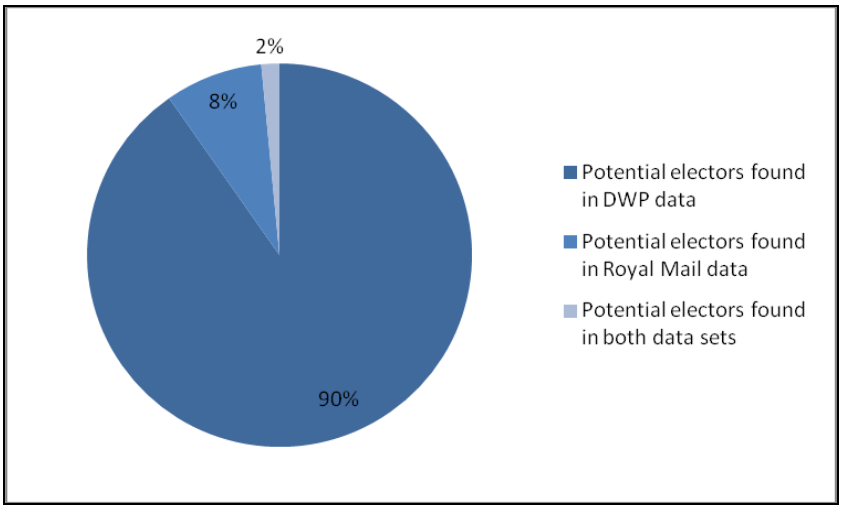
Whilst the entire register for each area was matched against the Royal Mail data for the purposes of identifying individuals who had moved within area, for the DWP data matching was only undertaken on those areas that the participating authorities intended to follow-up as part of the pilot. Five of the 14 areas requested for their full register to be matched against DWP data. Across these five areas the volume of

[25] Whilst South Ribble's register was matched against DWP and Royal Mail data for the purposes of the pilot it should be noted that, owing to the delays in sending the data to EROs, South Ribble were unable to conduct any follow-up work using this data and therefore are not included in sections 4.2/4.3.

records returned by DWP ranged from six per cent in Wigan to 25 per cent in Tower Hamlets. Both Wigan and Tower Hamlets also participated in the Confirmation data matching pilots and a similar difference was observed in match rates in these pilots, for example whilst 83 per cent of Wigan's pre-canvass electoral register could be successfully matched against DWP data, the figure in Tower Hamlets was much lower at 55 per cent as Tower Hamlets has a relatively high rate of population mobility[26].

Figure 4.1a illustrates the proportion of records found in each data set for the five pilot areas that matched their full register against both Royal Mail and DWP data. It highlights that the majority of records originate from DWP[27] and that only a small minority of records were located on both the Royal Mail and DWP data sets[28].

Fig 4.1a: Proportion of records found in each data set for the five pilot areas that matched their full register against both Royal Mail and DWP data



Attainers and Students

Across the ten English/Welsh authorities the volume of records returned from the DWP data, which included all non-matched records for individuals aged between 16 and 18 years old (inclusive), ranged from between 3.1 per cent of the electoral register in Rushmoor to 5.6 per cent of the register in Wolverhampton (average 4.5 per cent)[29].

[26] https://www.gov.uk/government/publications/simplifying-the-transition-to-individual-electoral-registration

[27] As highlighted earlier this may be expected as Royal Mail data will cover only a relatively small proportion of home movers whom opt to pay for a redirection service and for whom Royal Mail have the appropriate permissions to share their data.

[28] It should be noted that the figures detailed here relate to the overall returns provided., In many cases they are larger than the sample sizes presented in tables in later sections as some areas chose to work off a smaller sample – e,g by selecting particular wards or taking a random sample of records and only following these records up.

[29] It is important to note that as the returns included all individuals aged between 16 and 18 years old inclusive some individuals within the data would be either too young or too old to be classed as an attainer (attainer status is calculated according to the publication date of the register, and whether an individual is under 18 years but will reach 18 before that date).

Across the eight English local authorities whose registers were matched against DfE data the volume of records returned ranged from between 0.5 per cent of the electoral register in Rushmoor to 2.0 per cent of the register in Wolverhampton (average 1.4 per cent)[30].

Across four areas whose registers were matched against data from the Student Loans Company the volume of records returned ranged from between 1.0 per cent of the electoral register in Wolverhampton to 1.8 per cent of the register in Tower Hamlets (average 1.5 per cent)[31].

[30] Welsh authorities are excluded from this figure as the Welsh Education data only included postcodes and not full addresses and therefore was not usable for the purposes of the pilots.

[31] Data excludes Ceredigion as the overall volumes of data were unusually low (less than 0.05% of the register) indicating an issue with the data that we were unable to resolve prior to publication.

Fig 4.1b: Attainer data - proportion of records found in each data set for the eight English local authorities whose data was matched against DfE and DWP data

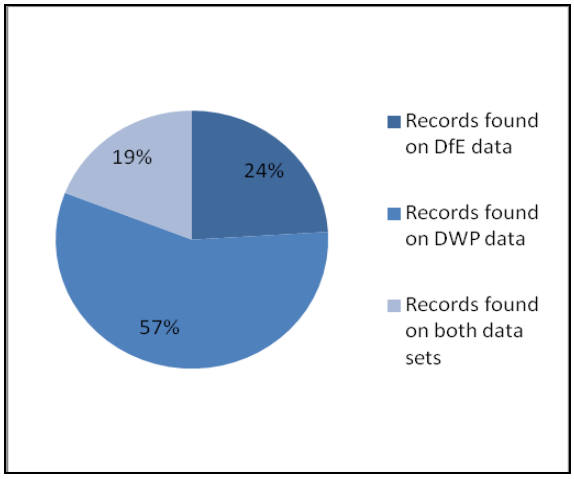
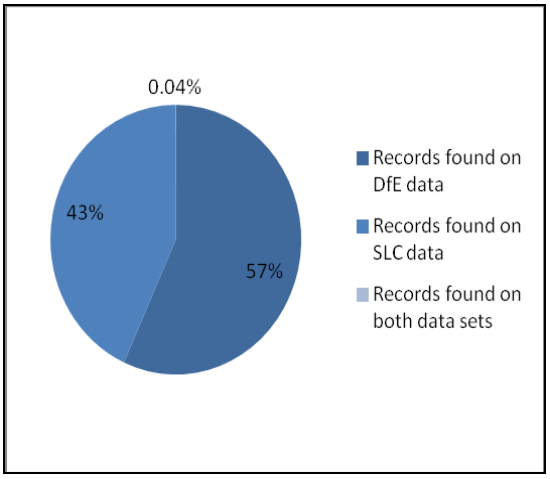


Fig 4.1c: Student data - proportion of records found in each data set for the four English local authorities whose data was matched against DfE and SLC data



Figures 4.1b and 4.1c illustrate the proportion of records found in each data set among pilot areas that matched against DfE/DWP data and DfE/SLC data for the attainers and students options respectively. Figure 4.1b illustrates that whilst the majority of records in the attainers data were located within DWP data there is significant overlap between the records found on DWP and DfE data sets[32].

In contrast Figure 4.1c shows that there is almost no overlap between records found on SLC data and DfE data, which can be expected given that they cover different tiers of education (higher education and secondary education respectively).

4.2 Results of checking data returns against locally held data

Reflecting some of the issues identified in the previous section, guidance provided to participants in the pilot recommended that all areas conduct additional checks of the data provided against their local register. Participants were also invited to check their data against other locally held data sets where they had the capacity and capability to do so.

Checks against the electoral register

The vast majority of areas did conduct these checks prior to writing out to individuals. Table 4a provides a summary of the results of participating areas checks of the data against their local registers across each of the options for data matching against national data sets. It illustrates that, on average, between 11 and 17 per cent of the records returned to EROs were found to be on the register when manual checks were undertaken.

[32] DWP data included individuals from 16 years old up to the age of 18 therefore may be expected to cover a greater volume of individuals than schools data which will include only school age individuals.

Table 4a: Proportion of records checked against the electoral register that were found to already be registered

	Student data (4 areas)[1]	Attainers data (8 areas)[2]	Home-movers data (12 areas)[3]
Average	17%	11%	15%
Min	8%	1%	4%
Max	29%	23%	27%

Notes:1. Excludes data from Ceredigion due to incomparable source data. 2. Excludes Pembrokeshire & South Ribble who did not conduct checks/provide data on checks. 3. Excludes Southwark & Pembrokeshire who adopted different sampling approaches to their data checks meaning their data is not comparable to other areas.

As discussed in more detail in the previous chapter, limitations in the automated matching algorithm mean that in some cases records are inaccurately identified as not matching to the electoral register due to differences in format and/or spelling of the names and addresses between data sets. Whilst refinements to the matching process/algorithms are likely to reduce the volume of these cases, it will not be possible to prevent this entirely.

Checks against other locally held data sources

A small number of pilot areas also conducted matching against other local data sets such as Council Tax and Housing Benefit records. The approach to this matching (e.g. records selected for matching) and the data sets used for this varied between areas, meaning that it is not possible to draw robust conclusions from the data. However, the data available indicates that within the records returned to EROs from national data sets a significant proportion of individuals could also be located within locally held data sets, ranging from 1% to 46% of attainers records checked in five areas[33] and from 26% to 88% of home-movers records checked in six areas[34]. Comparable data for student records is not available.

Whilst these figures should be interpreted with caution, the available results indicate that home-movers records were more likely to be found in local data than attainers. This may be expected as the most commonly used data sources for local matching in electoral registration are Council Tax and Housing Benefit records which may be less likely to include attainers.

[33] Based on data provided by the following pilot areas: Ceredigion, Powys, Sunderland, Tower Hamlets and Wolverhampton

[34] Based on data provided by the following pilot areas: Ceredigion, Greenwich, Powys, Sunderland, Tower Hamlets and Wrexham

Overall, these results further emphasise the value of checking records against the electoral register and where possible other local data sets. However, as discussed in more detail in the previous chapter, pilot areas also reported that this element of the data matching process was the most resource intensive part of the process, with the vast majority of pilot areas undertaking this process manually. Pilots reported that this was an extremely time consuming task and expressed concerns about their ability to resource this in a business as usual scenario.

4.3 Results of the follow-up activity

As mentioned earlier, the primary legislation which gives power to set up data schemes requires them to be based on proposals made by local registration officers. However, to ensure consistency for the purposes of evaluation all areas were asked to follow the same broad process for following up potential electors identified in the data matching. This included conducting at least one write-out to potential new electors identified during the data matching. Areas could then choose to conduct any additional reminder write-outs or canvassing dependent on their capacity.

The full results of the follow-up work are provided in Annex A and are summarised in tables 4.3a-e below. Across the results from all options a wide variation in results is reported and there are a range of factors that are likely to have contributed to this variation which it is important to consider when interpreting the results. These include:

- It is known that rates of registration vary by demographic groups and therefore differences in the overall populations served by participating areas may have impacted on response rates[35]. In addition, as described in the methodology section, participants could choose whether to conduct data matching across the entire register or whether to select specific geographical wards within their area, meaning that some areas concentrated their efforts on areas that have historically had relatively low registration rates.
- Differences in approach to processing the data may also impact on response rates. Where areas did not have the capacity to follow up the complete sample of records returned to them participants were asked to either follow up all records found within a limited number of wards or to select a random sample of records from across the data sets. Whilst the majority of areas broadly followed this approach, some areas excluded records based on other factors such as addressing issues or simply running out of time to complete the checks of the data within the timeframe given[36].
- In addition, some areas adopted differing approaches to the write-out as well as any subsequent follow-up activity which may have impacted on response levels. For example, where areas sent out additional reminders or included additional incentives to respond this may have enhanced the response level.

[35] Great Britain's Electoral Registers 2011', Electoral Commission, 2011.

[36] As noted previously the timescales for processing and following up the data were shortened owing to delays in getting the legal agreements required for data sharing in place.

- Linked to the point above, differences in the wording used in the invite letter may also impact on response levels. However, participants were provided with a template letter, and feedback from the pilots indicated that the majority of areas used this letter with only minor amendments: therefore the impact of this is expected to be relatively small.

To facilitate the interpretation of results, key points on the approach adopted for each individual area are provided within the following results tables.

Home movers

Table 4.3a below shows that, overall there was wide variation in the registration rates achieved. The volume of registrations as a proportion of potential electors written out to ranged from 2.9 per cent to 33.5 per cent overall.. The registration rates for DWP records ranged from 3.3 per cent to 25.7per cent whilst the range for Royal Mail records was 2.3 per cent to 24.2 per cent.

One of the recommendations arising from our evaluation of the 2011 pilots was that any future data matching should match to records which have been updated or had some activity within the previous 3, 6 or 12 months and that a record date should be provided where possible. In the 2012 pilots the DWP data was provided to EROs with a marker indicating whether the data had been updated in the past 0-3 months, 3-6 months, 6-9 months or 9-12 months. Participants in the pilots reported that the inclusion of this currency marker was useful and a number of areas were able to provide data on the results of their follow up (see Table 4.3a). Overall, the feedback from all pilot areas indicates that limiting the currency of the home-movers data to records which have been updated in the last 6 months would be beneficial. This would have the added benefit of reducing the overall volume of records returned to EROs (and consequently the time taken to complete checks of their data). Analysis of the DWP data indicates that this would reduce the overall volume of records by just over one third (see Figure 4.3a)

Figure 4.3a: Currency of all DWP home-mover records returned across pilot areas

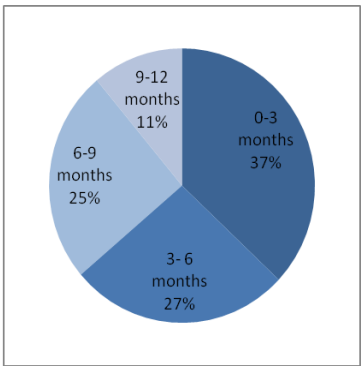


Table 4.3a: Registration rates achieved by currency of data

	0-3 months	3-6 months	6-9 months	9-12 months
Greenwich	10%	3%	2%	2%
Lothian	9%	5%	7%	2%
Powys	42%	28%	13%	11%
Wrexham	13%	8%	4%	3%

Notes: Data is only available for four pilot areas and the overall sample sizes are small therefore figures should be treated with caution and are illustrative only.

Table 4.3a: Home-movers data – potential electors contacted and registrations achieved by pilot area and data source

Local Authority	Approach to pilots and sampling	Sample size	Database	Number of potential new electors written to	New electors registered	% letter registration rate	Number of potential new electors selected for personal canvassing	New electors registered	% canvass registration rate	Total new electors registered	% registration rate overall
Ceredigion	Were interested in matching 3 wards.	428	**DWP**	282	29	10.3%				29	10.3%
			Royal Mail	34	8	23.5%				8	23.5%
			Combined	5	1	20.0%				1	20.0%
			total	321	38	11.8%				38	11.8%
Conwy	Were interested in 3 areas, covering around a third of their total electorate.	3858	**DWP**	2372	333	14.0%				333	14.0%
			Royal Mail	616	149	24.2%				149	24.2%
			Combined	66	24	36.4%				24	36.4%
			total	3054	506	16.6%				506	16.6%
Greenwich	They took a random sample of 3148 DWP records (ensuring a spread over time for the currency of the records) and 290 Royal Mail records.	3438	**DWP**	1623	132	8.1%				132	8.1%
			Royal Mail	129	18	14.0%				18	14.0%
			total	1752	150	8.6%				150	8.6%
Lothian	Were interested in matching to 2 areas of register covering approximately a quarter of their electorate. Decided not to use any Royal Mail data due to inclusion of individuals below age of registration in the records so only DWP results are reported.	453	**DWP**	373	30	8.0%				30	8.0%

Local Autho-rity	Approach to pilots and sampling	Sample size	Data-base	Number of potential new electors written to	New electors registered	% letter registration rate	Number of potential new electors selected for personal canvassing	New electors registered	% canvass registration rate	Total new electors registered	% registra-tion rate overall
Pembro-keshire	Removed a large number of records with addressing issues (mostly DWP records). Printed out remaining records and took a random sample of five from each page. Sample size is not directly comparable to other areas since they removed records which they found on the register at this stage.. Sampling approach could mean that records with better quality addresses were chosen, which could in turn affect response rates. Results for DWP and Royal Mail are reported on jointly.	540	**DWP and Royal Mail**	540	181	33.5%				181	33.5%
Powys	Were interested in matching to six wards covering around 10% of their electorate. They didn't check all of the records, so 330 weren't checked against the register before they wrote out to them due to time constraints.	980	**DWP**	771	198	25.7%	85	25	29.4%	223	26.1%
			Royal Mail	110	21	19.1%	2	0	n/a	21	18.8%
			Com bi-ned	43	27	62.8%	0	0	n/a	27	62.8%
			total	924	246	26.6%	87	25	28.7%	271	26.8%
Renfrew-shire	Excluded those that they found on the register and then randomly selected approx 20% to follow up. The sample for DWP has therefore been calculated as 610 as they wrote to 535 of the sample (after excluding those already registered) which is 88% of the sample (this is to make the results comparable).	1011	**DWP**	535	57	10.7%	75	2	2.7%	59	9.7%
			Royal Mail	294	66	22.4%	70	7	10.0%	73	20.1%
			total	829	123	14.8%	145	9	6.2%	132	13.6%

Local Authority	Approach to pilots and sampling	Sample size	Data-base	Number of potential new electors written to	New electors registered	% letter registr-ation rate	Number of potential new electors selected for personal canvassing	New electors regis-tered	% canvass regist-ration rate	Total new electors regis-tered	% regist-ration rate overall
	Interested in mining two wards (approximately 7000 electors for DWP) and the whole borough for Royal Mail. Sent a reminder letter to those outstanding on DWP after two weeks (351 records) and a reminder to Royal Mail for those outstanding for two postcodes for DWP data (393 records). One week later they had 291 electors outstanding from DWP data which they checked against council tax and archives and reduced down to 176 who they then canvassed over one weekend, in one postcode all were knocked twice and in the other once. The results of the write out for this pilot should therefore be expected to be slightly higher due to the reminder letters.		**DWP**	475	78	16.4%	176	24	13.6%	102	15.7%
			Royal Mail	2842	210	7.4%	0	0	n/a	210	7.4%
Richmond		4025	**total**	3317	288	8.7%	176	24	13.6%	312	8.9%
	Interested in 3 wards with an electorate of approximately 33000.		**DWP**	4661	156	3.3%				156	3.3%
			Royal Mail	4348	101	2.3%				101	2.3%
			Combi ned	68	5	7.4%				5	7.4%
Southwark		9145	**total**	9077	262	2.9%				262	2.9%
	Matched the whole register and checked all potential electors against local data (council tax and housing benefit data).		**DWP**	12356	946	7.7%				946	7.7%
			Royal Mail	492	42	8.5%				42	8.5%
			Combi ned	364	121	33.2%				121	33.2%
Sunderland		13799	**total**	13212	1109	8.4%				1109	8.4%

Local Authority	Approach to pilots and sampling	Sample size	Data-base	No.of potential new electors written to	New electors registered	% letter registra-tion rate	Number of potential new electors selected for personal canvassing	New electors registered	% canvass registration rate	Total new elec-tors regis tered	% registration rate overall
Tower Hamlets	In the time allowed they managed to check all 0-3 month currency and part of 3-6 month currency. Only managed to check a total of 21% of the returned data due to lack of time. A total of 530 names were deleted from the register as a result of checks undertaken during the pilot.	9317	**DWP**	8183	714	8.7%				499	6.1%
Wrexham	Checks against local data were only carried out for non-responders to the follow up. Of the 1741 non responders 456 could be found on council tax.	2265	**DWP**	2035	190	9.3%				190	9.3%
			Royal Mail	45	5	11.1%				5	11.1%
			Comb-ined	16	0	0.0%				0	0.0%
			total	2096	195	9.3%				195	9.3%
Wigan	The results reported here are for both Royal Mail and DWP as they couldn't differentiate between the forms returned.	13539	**DWP**	12133	1752	13.0%				1752	13.0%
			Royal Mail	1065							
			Comb-ined	304							
			Total	13502	1752	13.0%				1752	13.0%
TOTAL		**49259**		**43968**	**4485**	**10.2%**	**408**	**58**	**14.2%**	**4543**	**10.2%**

Notes: Owing to different approaches adopted by pilot areas total figures should be interpreted with caution.

Attainers

Table 4.3b and c detail the registration rates achieved for the attainers pilots. Table 4.3 b shows that registration rates for attainers written to after being identified by DWP records ranged from 12.7 per cent in Rushmoor to 17.5 per cent in Powys (rising to 17.8% when including a personal canvass). The registration rates for those found on both data sets were slightly higher, ranging from 12.7 per cent in Sunderland to 40.3 per cent in Powys (dropping to 39.2% when the personal canvass is included), but it should be noted

that the sample sizes were smaller. Results for Harrow should be treated with caution since they only wrote to people found on DfE data and not elsewhere, which may account for the lower response rate.

Table 4.3c shows the results for Coventry and Wolverhampton who reported on the results of people found on either database together. The registration rates ranged from 11.1 per cent to 18.5 per cent and when including a personal canvass rose to 36.2 per cent. The results suggest that DWP data might be more useful but where individuals were found on both it might help improve registration rates.

Table 4.3b: Attainers data – potential electors contacted and registrations achieved by pilot area and individual data sources

Local Authority	Approach to pilots and sampling	Total electors on whole register	Sample size	Data-base	Number of potential new electors written to	New electors registered	% letter registra-tion rate	Number of potential new electors selected for personal canvassing	New electors registered	% canvass registra-tion rate	Total new electors registered	% registra-tion rate overall
Ceredigion	They went on to exclude a large number of files from Welsh data where only the postcode and no address was available. There were a total of 1620 records omitted as the data was deemed unusable of these 603 from were from DfE/Welsh education, 972 from DWP and 45 found on both data sets.			**DWP**	594	83	14.0%				83	14.0%
				Combined	17	4	23.5%				4	23.5%
		59356	2368	**total**	611	87	14.2%				87	14.2%

Local Authority	Approach to pilots and sampling	Total electors on whole register	Sample size	Data-base	Number of potential new electors written to	New electors registered	% letter regist-ration rate	Number of potential new electors selected for personal canvassing	New electors registered	% canvass regist-ration rate	Total new electors registered	% registr-ation rate overall
Harrow	There was some confusion over the one of the column headings on the data return and as a result the pilot did not use the data returned by DWP or those who were found on both DWP and DfE records. Therefore only DfE data was used, this should be taken into consideration when viewing the response rate.	178266	584	**DfE**	519	13	2.5%				13	2.5%
Powys	They excluded 14866 records from Welsh data which had no addresses just postcode. Of the 3240 DWP records 259 had a date of birth outside of the attainer range (too young, they kept in those who were too old), others were outside of the target area, leaving 391.	103205	391	**DWP**	171	30	17.5%	14	3	21.4%	33	17.8%
				Comb-ined	206	83	40.3%	16	4	25.0%	87	39.2%
				Total	377	113	30.0%	30	7	23.3%	120	29.5%

Local Authority	Approach to pilots and sampling	total electors on whole register	Sample size	Data-base	Number of potential new electors written to	New electors registered	% letter registration rate	Number of potential new electors selected for personal canvassing	New electors registered	% canvass registration rate	Total new electors registered	% registration rate overall
Rushmoor	received 3090 from national data, 1319 removed as 1117 already 18, 176 were too young, leaving 1771 of attainer age to be checked. As well as a letter they also produced a flyer which explained the process of registration to potential electors and offered a small incentive - a chance to win a £20 Cineworld voucher	66104	1771	**DWP**	597	76	12.7%				61	10.2%
Sunderland				**DWP**	3920	227	5.8%				227	5.8%
				DfE	341	18	5.3%				18	5.3%
				Combined	1878	239	12.7%				239	12.7%
		219225	6463	**Total**	6139	484	7.9%				484	7.9%
TOTAL		626156	5114	Total	8243	773	9.4%	30	7	23.3%	765	9.2%

Notes:1) Tower Hamlets also conducted the attainer data matching option but their results have not been presented here as they were unable to provide separate data on registrations for students and attainers. They did however, register a total of 215 people from data identifying students and attainers.2) Rushmoor have reported that since the data collection period ended they have received a further 40 registrations. 3) Owing to different approaches adopted by pilot areas total figures should be interpreted with caution.

Table 4.3c: Attainers data – potential electors contacted and registrations achieved by pilot area and combined data sources

Local Authority	Approach to pilots and sampling	total electors on whole register	Sample size	Data-base	Number of potential new electors written to	New electors registered	% letter registration rate	Number of potential new electors selected for personal canvassing	New electors registered	Total new electors registered	% canvass registration rate	% registration rate overall
Coventry	Of the total records, 2881 were over 18 (they did write to these people and registered 130 of them), 485 were under 16 and so have also been excluded from the sample (and were not written to). There were 28 with 'bad addresses' which have been included in the area. The results have been reported on together.	237547	5360	**DWP and DfE**	5280	588	11.1%			588		11.1%
Wolverhampton	They excluded duplicate records between the student and attainer files (were working on both options), leaving them with a sample of 4178. For personal canvassing they selected 3 of 20 wards, the canvassers delivered letters and registered them at the door so not a 2nd stage to already having received a letter. They looked at the data by ward and didn't get to 5 of the 20 wards.	177558	4178	**DWP and DfE**	2011	372	18.5%	1141	409	593	19.4 %	36.2%
Total		415105	9538	**Total**	7291	960	13.2%	1141	409	1181	19.4 %	36.2%

since the data collection period has ended Wolverhampton have reported that they have received a further 85 registrations. It is not possible to say whether these are students or attainers or from which data base they were first identified.

Students

Tables 4.3 d and 4.3e show results for the four authorities seeking to identify students. Ceredigion did not use any Welsh data due to the lack of address information and did not register any students from SLC data. Greenwich had an overall response rate of 9.9 per cent, with better responses to people found on DfE data than SLC. Coventry and Wolverhampton reported on the results of the two datasets together and their response rates ranged from 5.4 per cent to 18.5 per cent.

Table 4.3d: Students data – potential electors contacted and registrations achieved by pilot area and individual data sources

Local Authority	total electors on whole register	Sample size	Caveats re data/approach taken	Database	Potential new electors identified	Number of potential new electors written to	New electors registered	% letter registration rate
Ceredigion	59356	704	Did not use any Welsh education data due to lack of addresses so not reported on here.	**SLC**	5	4	0	0.0%
Greenwich	170966	1194	776 records had incomplete addresses. Of the DfE records 2606 were too young so were excluded from working sample. Using date of birth random samples taken across age ranges to produce sample size.	**DfE or Welsh Education**	231	314	41	13.1%
				SLC	437	425	32	7.5%
				total	668	739	73	9.9%
TOTAL	230322	1898			673	743	73	9.8%

Notes: Tower Hamlets also conducted the attainer data matching option but their results have not been presented here as they were unable to provide separate data on registrations for students and attainers. They did however, register a total of 215 people from data identifying students and attainers.

Table 4.3e: Students data – potential electors contacted and registrations achieved by pilot area and combined data sources

Local Authority	total electors on whole register	Sample size	Caveats re data/approach taken	Database	Number of potential new electors written to	New electors registered	% letter registration rate (new electors as proportion of letters sent)
Coventry	237547	6939	Of the 6939 records, 24 had addresses outside of Coventry and 159 had bad addresses so they were unable to write to them.	**DfE and SLC combined**	6195	337	5.4%
Wolverhampton	178589	2945	They excluded duplicate records between the student and attainer files (were working on both options).	**DfE and SLC combined**	1456	270	18.5%
TOTAL	815328	25408			7651	607	7.9%

since the data collection period has ended Wolverhampton have reported that they have received a further 85 registrations. It is not possible to say whether these are students or attainers or from which data base they were first identified.

Chapter 5
Summary and conclusions

This report presents the findings from the 2012 data matching pilots which were set up to test the potential effectiveness of using data matching, whereby records on the electoral register are matched against other sources of public data to improve the completeness and accuracy of the electoral register.

The pilots were designed to target specific groups that are traditionally under-registered, namely recent home-movers, attainers and students. By matching the electoral registers of participating authorities against other trusted data sources, the aim of the matching was to identify individuals who are not currently registered and provide these details to EROs, enabling them to contact the individual and invite them to register.

5.1 Key findings – matching against national data sets

Effectiveness of data sources

Across the data sources there was a wide variation in the registration rates achieved between pilot areas. There are a number of factors related to the different approaches areas took to the selection of the sample of data to follow up that are likely to explain a part of this variation. These findings should therefore be interpreted with caution. (See Chapter 4 for full details)

Data matching to identify attainers

In order to identify unregistered attainers, electoral registers were matched against DWP data and national education data. Across the ten authorities whose data was matched against DWP data, the volume of records (potential new electors) returned ranged from an equivalent of between 3.1 to 5.6 per cent of the electoral register (average 4.5 per cent). Across the eight local authorities whose data was matched against DfE data the volume of records returned ranged from between 0.5 per cent to 2.0 per cent of the electoral register (average 1.4 per cent). This demonstrates that the data matching process was able to identify potential new electors, although it is important to note that the pilots also showed that not all of these records were potential new electors and included records of individuals who were already

registered, ineligible to register or no longer present at the address, as discussed later in the chapter[37].

Of those records subsequently followed up by the pilots[38], amongst those areas that provided data per individual data source, the pilots showed that registration rates for attainers written to after being identified from DWP records ranged from 12.7 per cent in Rushmoor to 17.5 per cent in Powys (rising to 17.8% when including a personal canvass). The registration rates for those found on both DfE and DWP data sets were slightly higher, ranging from 12.7 per cent in Sunderland to 40.3 per cent in Powys (dropping to 39.2% when the personal canvass is included), but it should be noted that the sample sizes were smaller. The results for the two pilot areas who reported on the results of people found on either database together showed that registration rates ranged from 11.1 per cent to 18.5 per cent and when including a personal canvass rose to 36.2 per cent.

These results suggest that DWP data might be more useful for attainers, but where individuals were found on both it might help improve registration rates.

Pilots also fed back that accessing data on attainers was seen as particularly beneficial as they are less likely to be able to identify this group in the local data sets that they currently use for matching. (It should however be noted that EROs in unitary authorities are currently permitted to access local education data for the purposes of maintaining the register, although it is not commonly used).

Data matching to identify students

In order to identify potentially unregistered students, electoral registers were matched against education data (as with attainers) and data held by the Student Loans Company.

Across the four areas whose data was matched against SLC data the volume of records returned ranged from between 1.0 per cent to 1.8 per cent of the electoral register[39]. Of those records subsequently followed up[40],Ceredigion did not register any students from SLC data, Greenwich had an overall response rate of 9.9 per cent, with better responses from people found on DfE data than those found on SLC data, and Coventry and Wolverhampton reported on the results of the two datasets together, (their response rates ranged from 5.4% to 18.5%). The DfE data sets used were in practice the same as those used for attainers, therefore, and there is limited evidence to support the use of data matching to identify students as a discrete group.

37.

38 It should be noted that the majority of areas did not follow up all records and pilot areas adopted different approaches to selecting the records followed-up.

39 As with attainers this demonstrates that data matching was able to identify potential new electors although it should be noted that the pilots also showed that not all of these records were potential new electors and included records of individuals who were already registered, ineligible to register or no longer present at the address, as discussed later in the chapter.

40 It should be noted that the majority of areas did not follow up all records and pilot areas adopted different approaches to selecting the records followed-up.

Data matching to identify home-movers

In order to identify potential new electors amongst recent home-movers the electoral registers were matched against data held by Royal Mail and DWP. Across the 14 areas participating in the home-movers pilot, the volume of records returned by Royal Mail equated to between 0.6 to three per cent of the their total register size (average 1.5 per cent). Across the five areas whose full register was matched against DWP data the volume of records returned ranged from six per cent to 25 per cent[41].

Amongst those records subsequently followed up by the pilots[42] the volume of registrations (as a proportion of potential electors written out to) ranged from 2.9 per cent to 33.5 per cent overall.. The registration rates specifically for DWP records ranged from 3.3 per cent to 25.7 per cent, whilst the range for Royal Mail records was 2.3 per cent to 24.2 per cent.

A number of specific benefits were identified with the data for home-movers, including the relative address quality of the Royal Mail data and the availability of UPRNs within DWP data which will have facilitated the matching process and data checking. In addition, evidence suggests that by limiting the currency of the data to records updated within the last 6 months it may be possible to produce a more targeted data set and therefore increase registration rates.

However, it is also worth noting that evidence from the pilots suggested that of the options tested, home-movers data had the greatest overlap with local data indicating that some of the potential electors contacted could have been identified by data sets that EROs are already using. In addition there were some particular issues with the data, specifically that the Royal Mail data available in the pilots did not include date of birth and therefore included individuals below the age of registration which would need to be excluded from any future exercise[43]. In addition, feedback from pilots showed that in some cases the addresses provided were forwarding addresses only, meaning that the individual did not or never had resided there. Finally, as Royal Mail is a state-owned public limited company the cost of accessing their data was higher[44].

Process Evaluation

The evaluation has also highlighted a number of issues with the process of data matching that would need to be considered and/or resolved if data matching were to be rolled out. Some of these relate to the practical application of data matching,

---

[41] As with attainers and students this demonstrates that data matching was able to identify potential new electors although it should be noted that the pilots also showed that not all of these records were potential new electors and included records of individuals who were already registered, ineligible to register or no longer present at the address, as discussed later in the chapter.

[42] It should be noted that the majority of areas did not follow up all records and pilot areas adopted different approaches to selecting the records followed-up.

[43] Royal Mail have subsequently informed us that they are able to exclude individuals below the age of registration.

[44] This is because Royal Mail charged for their data which is routinely sold to outside organisations, whilst the other DHOs did not. In addition the processing was carried out by their partner 'Transactis' which incurred additional cost

including suggestions as to how the process may be improved for any future exercises, whilst others relate to the more inherent challenges of using in this way data that was originally collected for a purpose other than that of issuing invitations to register to vote.

The key issues identified are summarised below.

Data Quality

A number of issues with the quality of data when used for the purposes of identifying potential electors were observed. It is important to note that not all of these issues are reflective of the quality of the source data, which is collected for a different use. However, some of the issues identified do indicate that there is scope for improvement in the quality of centrally held data sets.

- Duplicate data – whilst some de-duplication of data was conducted in the pilots, adding additional steps to further identify and remove duplicate entries both within and across options would be beneficial for any future exercise
- Address quality – with the exception of Royal Mail data the address quality of the data was reported to be poor. A potential option to resolve this may be to add an additional step to the matching process whereby the address information is matched to the property list held by each ERO, enabling the address information provided by the ERO to be returned to the ERO as part of the data. In addition, limiting the return data to include only records where the address can be matched to the ERO property list but where there is no identity match[45] could be one way to provide a more targeted data set. This has the potential to make the follow-up process for EROs simpler by requiring less data cleansing and may improve registration rates, although it may result in the exclusion of some genuine potential new electors.
- Ineligible electors – the source data sets will inevitably include individuals who are ineligible to register (e.g. because of their nationality). This emphasises the need for EROs to conduct additional checks on the data returned to them prior to writing out to individuals. However, it will not be possible completely to prevent invitations to register being sent to individuals who are not eligible to register.
- Other issues – the pilots highlighted that the data returned to EROs included some individuals who were deceased and some records of potentially sensitive addresses (e.g. Women's Refuges). Whilst steps are taken in the process to prevent this, it is not possible completely to remove the risk of this data being included, and in many cases locally held data (for example local registers of deaths) will include more accurate information. Guidance to EROs must be explicit about these limitations and what checks may be undertaken as mitigation.

Identification of data inaccuracies

[45] An identity match is a match of the name details of the record.

The follow-up processes undertaken by local authorities can help to identify inaccuracies in the electoral register, but also highlight inaccuracies in the other data sets used for matching, which raises some specific issues:

- Pilots reported that one of the most common reasons for enquiries from members of the public was that an individual had received a letter at their property for someone who did not live there. In some cases individuals were concerned about why the ERO had incorrect information, where it had come from, and what action could be taken to correct the inaccuracy. If future data matching exercises were to be conducted EROs are likely to require guidance on what they can advise individuals in these circumstances[46].
- If data matching was to be used on an ongoing basis, a mechanism would be required enabling the ERO to record this information and/or provide feedback to the DHOs regarding the inaccuracies identified, in order to prevent duplication of effort or individuals being contacted again after reporting an inaccuracy. Options will require feasibility testing, including consideration of the legal issues around storing such data, particularly where reports of inaccuracies have been made through a third party.
- Whilst the primary purpose of the data matching is to improve the completeness and accuracy of the register, there may be cases where the inaccuracies identified are indicative of non-electoral fraud. Clear guidance would need to be provided to EROs and DHOs on what action they are permitted and/or required to take in these cases.
- It is important to note that DHOs cannot guarantee the accuracy of their data sets and therefore information should be treated simply as one source of intelligence to be considered alongside other information. Guidance provided to EROs should be explicit about this and the known limitations of the relevant data sets. The extent of any issues arising from these limitations would also need to be closely monitored.

Resource requirements

- Pilots reported that the most resource intensive part of the process was the checking of the data against their local register and/or other locally held data sets, which for the vast majority was completed manually.
- Many areas raised concerns about the scalability of the process given the resource intensiveness of the data checks, and the majority of areas fed back that the process would need to be more automated in order to reduce the burden on their time.
- The pilots have identified a number of ways in which the process could be refined to improve the quality of the data returned and/or reduce the volume of records returned. There is also the potential for greater automation of the process (through integration with EROs' existing EMS systems) to be

[46] Whilst individuals may be able to contact a DHO directly to correct their own personal details this may not be possible where information is received from a third party.

beneficial in terms of the resources required to followup records. However, it is important to note that due to the inherent limitations of automated data matching some manual checking of records will always be required.

- The cost implications of delivering the system centrally will vary depending on the scope and scale of any future exercises, but they have the potential to be significant.
- In addition to any initial set-up costs, if data matching was to be used on an ongoing basis there are likely to be a number of potential ongoing central resource requirements including:
    - Co-ordination with DHOs. and funding relevant costs to the DHOs
    - Co-ordination with EROs
    - Matching the data and processing the files
    - Ensuring a legal basis for data exchange is in place and facilitation of secure transfer of data

Views on how data matching might work in the future

- Overall, the majority of the pilot areas report that having access to the data was valuable and that they would wish to use the data again given the opportunity. Participants cited the benefits of having access to a wider range of information, alongside local data, to assist in improving the completeness and accuracy of the register.
- However, it should be noted that many participants also reported that without improvements to the system that would reduce the burden of manually checking records they would have concerns about the practical feasibility of conducting such an exercise alongside their other work. In light of this, and given the overlap between records returned from the national data sets and local data sets, some areas questioned whether using local data may be more efficient.
- Given the associated resource implications for authorities undertaking data matching, if data was to be made accessible for this purpose, it is suggested that this should be voluntary, with EROs opting-in based on their own assessment of the relative value this type of data matching would have in their area and the groups and/or areas that they wish to target. However, predicting the likely demand for this is challenging, because whilst the majority of pilot areas expressed an interest in receiving this data they are a self-selecting sample, likely to be more enthusiastic about data matching, and cannot be considered representative of all areas.
- Data matching was identified as having the potential to be particularly beneficial as part of the Confirmation process. This is because individuals who provide their details in response to the invitation would effectively have already been confirmed as their details will have been data matched as part of the process. In addition, as some local areas will be planning to conduct local matching at this time anyway, there may be the possibility for

efficiencies to be achieved by using that matching both for confirming existing electors and identifying potential new ones.

- If data matching was to become part of business as usual, the preferred timing for the activity is likely to vary according to the groups targeted and the ERO's planned canvass activities. For example, pilots reported that students may most usefully be targeted at the start of the academic year in September or October. A flexible approach to the timing of the data matching may therefore be beneficial, although this would need to be Cabinet Officensidered against any related resource implications in terms of the management of the process.

5.2 Key findings – Two-Tier data matching

EROs have a statutory "right to inspect records kept in whatever form by the council by which [they have been] appointed". In unitary authorities EROs have access to all of the council's datasets, however, in a two-tier local government area a legal barrier means a lower tier authority holding the electoral register is unable to access the upper tier authority's data (which will include education records, amongst others).

As part of the current pilots the CO worked with four lower-tier authorities to test the usefulness and practicability of allowing two-tier local authorities to do internal data matching by removing this legal barrier. The aim of this piloting was to test the process and practicality of doing so.

Pilot areas reported some initial challenges in establishing the right contact within the Upper Tier authorities: however, once the relationships had been established they appeared to work well. Similar issues with the quality and limitations of the data, as experienced with the national data sets, were identified. However, a key benefit of using two-tier data (or other local data) is that ERO has direct contact with the data holder which can facilitate their understanding of the strengths and limitations of various data sets.

Overall participants were very positive about the perceived usefulness of being able to access the data[47]. However it should be noted that some participants raised concerns about the scalability in relation to the potential for upper-tier authorities receiving multiple requests from lower–tier authorities, suggesting that some level of co-ordination between authorities may be beneficial. In addition, in all areas the matching was undertaken by the lower-tier authorities and completed manually. Participants reported that being able to conduct the matching on an automated basis would be beneficial.

5.3 Conclusions

[47] The pilots were set up to test the process and practicability of data sharing therefore data on registration rates is not provided. However, some authorities did use the data to follow-up potential electors and where comparable data is available this is discussed amongst the results for attainers.

The data matching pilots have shown that providing access to national data may be beneficial in improving the completeness and accuracy of the register. However, the experience of the pilots shows that currently the process is extremely resource intensive, both for EROs and centrally, and unless the level of manual processing involved in the data matching could be reduced many areas reported that this burden would be prohibitive.

Our evaluation has identified a number of ways in which the system could be refined to address some of the issues identified. However, these options would need to explored and tested fully to ensure that the resource burden could be effectively reduced to a manageable level. Given the relatively low registration rates achieved in the pilots, the cost of implementing these changes and introducing a central system for data sharing would need to be carefully considered.

If future data matching exercises were to be undertaken, there may be particular value in carrying this out alongside Confirmation, as individuals registered through data matching have already been data matched (and therefore 'confirmed') as part of the process. However, it is suggested that any data matching exercise should be voluntary, with EROs opting in based on their own assessment of the relative value this type of data matching would have in their area and the groups and/or areas that they wish to target.

The pilots also identified an overlap between the records returned from national data sets and those found in local data sources. Supporting local authorities to use local data for the purposes of finding new electors may be valuable, either as a complementary activity or a potential alternative to using national data sets.

Finally, the pilots sought to test the feasibility of removing the legal barrier enabling lower tier authorities access to data which is held by the upper tier authority and which is currently available to all EROs in unitary authorities. The pilots showed that the lower tiers found data matching using this data a useful exercise. Whilst opening up the legal gateway to exchange data will not necessarily mean that all areas will opt to use this data, the pilots suggest there is potential value to this, particularly if support and/or guidance is provided to those authorities that wish to do so.