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- *Summary*

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

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.

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 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).

Studies have also shown that certain groups such as young people (including attainers[1]), 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).

In 2011 the Cabinet Office 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 Electoral Registration Officers (EROs) to contact these

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.

1 16/17 year olds who will become eligible to vote during the life of the electoral register.

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.

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

In addition, a separate aim of the pilots was to explore the potential value and practicability of opening the legal gateway to enable data to be shared between upper and lower tier authorities in two-tier local authorities[2].

This report provides a summary of the key findings from the 2012 pilots, which built on the learning from the 2011 pilots. The full evaluation report is available to download from gov.uk.

3. Overview of pilots

-

Participating areas

Electoral Registration Officers 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

[2] This would enable two-tier authorities access to data that is currently legally accessible by EROs within unitary authorities for the purposes of registration, for example locally held education data.

- **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]	Data matching Option(s) participated in	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

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 Cabinet Office 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 Cabinet Office 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[3]. Data included the names, dates of birth and postcodes of 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

[3] 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.

used by local authorities and therefore no results for this data set have been included in this evaluation report[4]. 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 Cabinet Office 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 Cabinet Office 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[5];

- 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;
- 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;

[4] 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.

[5] 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.

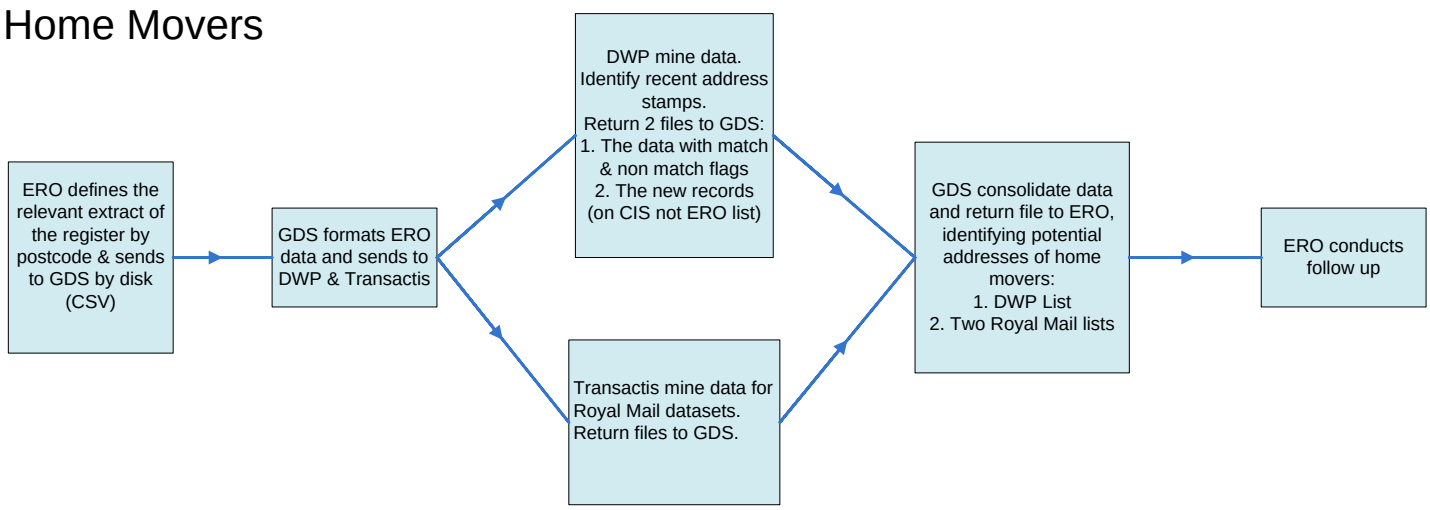
- Name, address and date of birth of individuals who are recorded in the Council's children's social care database as looked after by the Council and who are at least 16 years old but less than 19 years old; and
- Name, address and date of birth of individuals appearing in databases kept by the Council 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.

Data matching process

The primary legislation[6] 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.

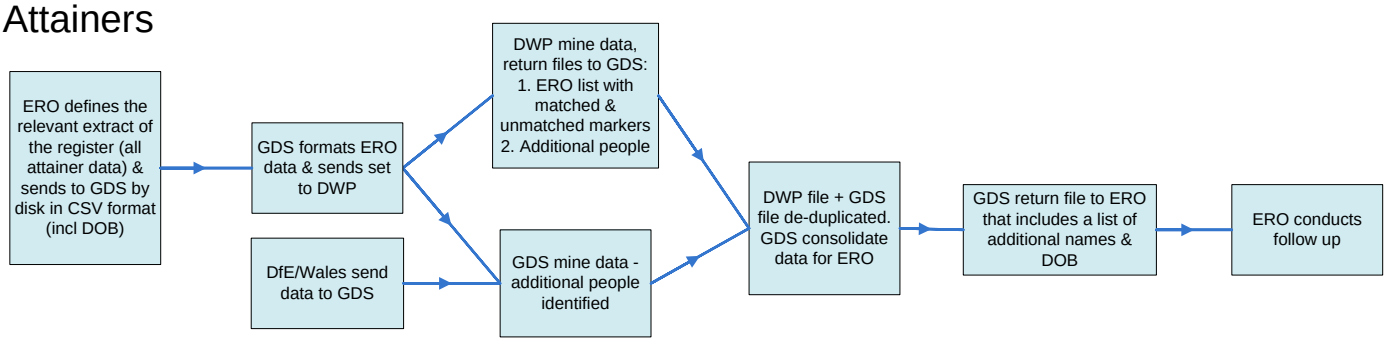
As the aim of the two-tier pilots was to test the feasibility of data matching and explore how the process might work, we did not set out a specific approach for matters such as the matching and transfer of data. Instead, we worked with the individual pilots to agree processes between the upper and lower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

Fig 1: Process maps for pilots using national data sources

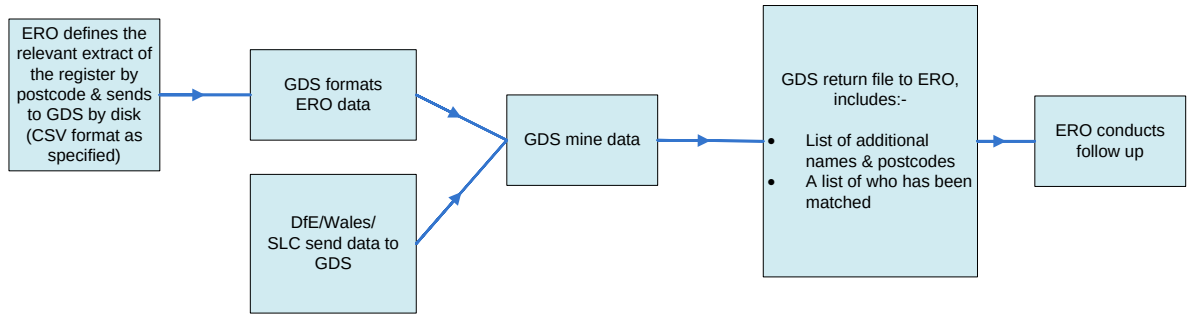


[6] The Political Parties and Elections Act 2009, section 36(1).

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[7]. The Cabinet Office 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)

[7] 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

- Full address (as it appeared in the original DHO data)
- 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.

4. Research aims

The key aims of the Cabinet Office evaluation 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).

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

- Cabinet Office led workshops which provided an opportunity for the Cabinet Office 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.

5. Key findings – matching against national data sets

5.1 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[8].

Of those records subsequently followed up by the pilots[9], 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[10]. Of those records subsequently followed up[11],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.

Data matching to identify home-movers

8.

9 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.

10 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.

11 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.

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[12].

Amongst those records subsequently followed up by the pilots[13] 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[14]. 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[15].

5.2 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, 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.

---

[12] 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.

[13] 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.

[14] Royal Mail have subsequently informed us that they are can exclude individuals below the age of registration.

[15] 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

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[16] 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

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

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

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[17].

- 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 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.

[17] 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.

- 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 considered against any related resource implications in terms of the management of the process.

6. 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[18]. 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.

7. Conclusions

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.

[18] 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.

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.