

Data mining pilot – evaluation report

July 2013

Translations and other formats

For information on obtaining this publication in another language or in a large-print or Braille version, please contact the Electoral Commission:

Tel: 020 7271 0500

Email: publications@electoralcommission.org.uk

Data mining pilot – evaluation report

July 2013

Contents

| | | |
|---|---|----|
| | Executive summary | 1 |
| 1 | Introduction | 11 |
| 2 | Set up – target groups and databases | 18 |
| 3 | Pilot processes – data transfer, matching, follow up | 27 |
| 4 | Costs | 38 |
| 5 | National data mining – judging success | 41 |
| 6 | County data mining – judging success | 57 |
| 7 | Recommendations | 60 |
| | Appendices | |
| | Appendix A – Department for Education | 63 |
| | Appendix B – Welsh Government, Department for Education and Skills | 68 |
| | Appendix C – Department for Work and Pensions | 69 |
| | Appendix D – Royal Mail | 76 |
| | Appendix E – Student Loans Company | 81 |
| | Appendix F – County databases | 84 |
| | Appendix G – National data mining, follow up work | 87 |
| | Appendix H – Pilot area profiles | 94 |

Executive summary

Background

- As part of the shift to individual electoral registration (IER) the UK government has been exploring the extent to which access to information held on national public databases can assist Electoral Registration Officers (EROs) in maintaining their electoral registers. This pilot, run by Cabinet Office, involved comparing registers with five national public databases as well as data held by four county councils in order to identify unregistered electors. This process is known as 'data mining'.
- The Electoral Commission has a statutory responsibility to report on the effectiveness of the pilot.
- Data mining was the subject of a previous pilot in 2011. We published our evaluation of that scheme in March 2012.¹

Set up – target groups and databases

- The pilot involved targeting three groups of electors known to have low registration rates – attainers (17 year olds and some 16 year olds), home movers and students.²
- Eighteen EROs were involved in the national pilot. They were provided with data held by the Department for Education (DfE), Welsh Department for Education and Skills (DfES), Department for Work and Pensions (DWP), Royal Mail and Student Loans Company (SLC) for the period of the pilot.
- Each of the EROs accessed data from at least two of the national databases.
- In addition, four district council EROs were given access to data held by their respective county council, with a specific focus on trying to identify unregistered attainers.³

¹ The Electoral Commission, *Data matching schemes to improve accuracy and completeness of the electoral registers – evaluation report* (March 2012)

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

² An attainer is an individual who will turn 18 before the end of a 12 month period starting from the next 1 December after the application is made i.e. if an application is made in spring 2013, the applicant will be eligible as an attainer if they turn 18 any time before December 2014.

Pilot processes – data transfer, matching, follow up

- There were considerable delays to the original timetable for establishing this pilot. A significant cause of the delays was the lack of capacity and resources within Cabinet Office (and the Government Digital Service (GDS), which is part of Cabinet Office) due to their workload related to the transition to IER.
- The delays affected the amount of work the EROs could do with the data.
- Three different organisations were involved in matching the national databases against the electoral registers in order to produce a list of potential unregistered electors. Each organisation used similar but not identical matching criteria. We have not been able to quantify any variation in the matching caused by differences in the matching criteria.
- For the national data mining, Cabinet Office's original intention was that pilot areas should adopt a fairly standardised approach to checking the data received and contacting the individuals identified, to ensure that results were comparable. In practice, however, the nature and extent of follow up work varied widely.
- Much of this variation was caused by practical difficulties, for example the need to spend more time than expected in ensuring the accuracy of the data received. However, some of the variation could have been avoided if there had been fewer delays and a greater level of support provided by Cabinet Office to pilot areas. In particular, a few areas told us they felt unsupported and were unclear about what to do.
- Although the wide variation in follow up activity has limited our evaluation in certain ways, we do have more complete and comparable data than in the 2011 pilot and so are able to reach more definite conclusions.

Costs

- It is not possible to produce an overall figure for the cost of this pilot. This is because we do not have final costs for all pilot areas or any costs for Cabinet Office (including GDS), who conducted much of the work.
- We are also therefore unable to estimate the cost per new elector registered or the likely cost of any national rollout. Any estimates of these would need to include the cost of coordinating and managing the pilot (the

³ Overall, 20 EROs participated in the pilot, with two participating in both the national and county data mining.

role taken by Cabinet Office in this pilot), as any future work with data mining would require some form of central coordination.

National data mining – judging success

- This evaluation sought to answer two key questions:
 - How effective is the data at identifying unregistered electors?
 - Is data mining a cost effective way of registering new electors?

How effective is the data at identifying unregistered electors?

- The evidence from this pilot suggests that data mining, as it was tested, is not a practical way of identifying unregistered electors.
- This is because, although the data returned to EROs did contain details of unregistered electors, it also contained significant numbers of existing electors, ineligible individuals and out of date information (where the individual was no longer resident at the given address).
- We have not made, and could not make, a consistent assessment of what proportion of records returned to EROs were in fact potential new electors. However, we have reported on what pilot areas found when they either checked the data against other data sources (for example their electoral registers) or when they received responses to their follow up work, as well as using their feedback on the processes and volume of work involved.
- The feedback from pilot areas was clear: **the amount of time and resources they spent on reducing the data provided to a list of likely new electors with usable address information was unsustainable and could not be incorporated into their ‘business as usual’ processes.**
- The reasons that so many existing electors and ineligible individuals were returned on the data include poor data specifications from Cabinet Office, currency restrictions not being tight enough and incomplete or poor quality addresses on some of the national databases.
- Inconsistent address formatting and incomplete addresses are likely to have contributed to the significant numbers of existing electors returned in the data (Cabinet Office could not provide the data which would have allowed for a definitive assessment). These problems also made it more difficult for EROs to use the information to write to the individuals identified.

- Some pilot areas were uncomfortable with the risk of writing to existing electors, individuals below the age of registering to vote or deceased people. Many areas spent a long time checking and cleansing the data before sending out registration forms.
- Some of these issues are likely to be relatively straightforward to resolve, for example creating better data specifications or improving data currency.
- However, some of the issues are more complex, particularly the issue of addresses being held in different formats between the national databases and electoral registers and, on some of the databases, being incomplete.
- The variable quality and formatting of addresses reflects the differing roles of addresses on the different databases. In many cases the address information is less important than it is on the electoral registers.
- Issues with addressing are likely to be only resolved in one of two ways: comprehensive addition of UPRNs to the national databases used for mining or extensive address standardisation at the initial stage of the mining process.⁴ Both of these solutions have resource implications.

Is data mining a cost effective way of registering new electors?

- The aim of this pilot was to see if data mining was an effective way of improving the completeness of the electoral registers, by identifying potential new electors who could subsequently be registered.
- In order to answer this question, we would need to assess the cost benefit of data mining by, for example, calculating the cost per new elector registered. However, we are unable to do this as Cabinet Office could not provide details of their expenditure on the pilot. As they managed the process and conducted much of the matching and data processing, their costs could be significant and are crucial in reaching any realistic assessment of cost effectiveness.
- In addition, we would need to assess the added value of providing EROs with access to national data compared to the data which they already have access to locally.
- We do know the numbers of new registrations achieved by pilot areas. These vary widely, from 2% to 31% of the sample the pilot area worked with, with an average of around 9%. In many cases, the actual numbers of new electors were low – on average around 300 individuals.

⁴ UPRNs are unique 12-digit codes assigned to each property at a local level. These local lists are then combined to form the National Land and Property Gazetteer in England and Wales and the One Scotland Gazetteer in Scotland.

- However, **the level of new registrations does not provide a clear assessment of the potential of data mining.** The registrations achieved in each area were partly the result of the approach taken to data checking and follow up work and this varied across the pilot areas. Also, in nearly all areas, the follow up work was limited to a single letter with no reminder or canvassing. The new registrations are therefore likely to be on the lower end of what could be achieved.
- In addition, the delays in the pilot timetable mean that not all of the registrations achieved could be reported here (as responses will have continued to come in up to and following publication).
- Part of the variation is also likely to be due to demographics. Some pilot areas had a higher response rate to their letters than others and in areas with higher levels of population mobility, it is likely that the data becomes out of date more quickly. From the data available, however, we cannot make any clear assessment of the impact of demographics on response rates.
- Taken on their own, some of the numbers of new registrations appear reasonable (given the general lack of personal canvassing in the pilot, we have taken a 10% registration rate as reasonable). However, **overall the numbers of new registrations are low in light of the time and resources spent to achieve them.** Feedback from the pilot areas supports this assessment.
- In order for data mining to be of practical use to EROs, the data returned would need to contain many fewer names of registered or ineligible individuals and have significantly improved address information. Our recommendations consider how that could be achieved.
- Finally, data mining would require a central organisation to take on the role of coordinating data transfer and processing the data – as Cabinet Office and GDS did for this pilot. Who that organisation would be (Cabinet Office, GDS or a different body), and what level of resource it would require, is a key question for any future use of data mining.

Individual databases – key findings

Department for Education (DfE)

- In terms of the usefulness of this data in identifying potential, unregistered electors:
 - The currency of this data appears to be good.
 - The addresses appeared to be more complete than those held in other national databases but a poor data specification from Cabinet Office meant that the format was inconsistent.
 - In the three areas that could separate their results, new registrations were an average of 9% of the sample of records they worked with.

- Overall, limited conclusions can be drawn about the DfE data as several pilot areas were unable to report their results separately for this database.

Welsh Department for Education and Skills (DfES)

- The DfES data does not include full addresses, only postcodes and as a result no pilot areas were able to do any follow up work with the data.
- This database is not suitable for the purposes of identifying unregistered electors because of the lack of address information.

Department for Work and Pensions (DWP)

- In terms of the usefulness of this data in identifying potential, unregistered electors:
 - There were numerous address issues including crucial address information being missing and extensive and confusing abbreviations.
 - A substantial number of the records returned related to existing electors.
 - The currency of the data was an issue. Evidence provided by some pilot areas indicates the data should be restricted to records which have been updated in the past three months.
 - Overall the number of new registrations achieved was 8%.
- We do not believe that any of the findings in this pilot call into question the use of DWP data for the confirmation process.⁵

Royal Mail

- In terms of the usefulness of this data in identifying potential, unregistered electors:
 - Some of the Royal Mail datasets requested by Cabinet Office included data relating to individuals below the age for registering to vote.
 - The addresses appeared to be more complete and consistent than those held in most of other national databases.
 - A substantial number of the records returned related to existing electors.
 - The currency of the data was an issue, with a substantial number of the records relating to individuals who were found to be no longer resident. This is likely to be partly a result of the two year update restriction placed on the data, which most pilot areas felt was too long.

⁵ Confirmation is the process by which existing electors will be matched against data held by DWP in order to retain them on the registers during the transition to IER. The findings from this pilot are not directly applicable to confirmation because data mining involves identifying and using non-matches between the registers and the DWP database and as such is inherently more likely to encounter issues with records held by DWP.

- Overall the number of new registrations achieved was 8%.
- Overall, Royal Mail data does not seem to be more effective than DWP data in identifying potential new electors, and there is a higher cost attached to using this data.

Student Loans Company (SLC)

- In terms of the usefulness of this data in identifying potential, unregistered electors:
 - There seemed to be issues with the addresses on this data being incomplete. Only one pilot area reported usable results for this database and they found that nearly a third of the addresses were quite clearly incomplete. SLC informed us that the addresses they provided to GDS were complete, so it seems that these issues may have arisen in the matching process, although we are unable to say for certain.
 - This pilot area reported a low number of new registrations.

County data mining – judging success

- This pilot did not try to assess the usefulness of access to county council data. Rather it was set up to provide qualitative feedback on the barriers and issues that would be faced if EROs in lower tier authorities tried to access the data held by an upper tier.
- There is therefore nothing in the findings to challenge the assumption that it would be sensible to equalise unitary authority and lower-tier authority EROs' access to data.
- However, it took a long time and a great deal of effort to establish the data sharing arrangements between the pilot areas and the county councils. It is clear that, if legal access was granted to lower tier EROs, they would still need to invest time in securing access to, and learning how to best utilise, the county data.
- Importantly, EROs would need to assess the cost effectiveness of accessing the data that became available to them – exactly as they should do with the local data they currently have access to.

Recommendations

National data mining

The findings from this pilot do not justify the national roll out of data mining. The concept of using national data to assist EROs may still have potential, but **data mining should not be implemented without further testing of the databases and processes.**

Data mining would require a central organisation to be responsible for managing the connection between national data holding organisations and undertaking data processing work. Cabinet Office undertook this role for this pilot.

The need for a central coordinating body is key as some data holding organisations, such as DfE, do not conduct the matching process themselves. However, the requirement is wider than this and includes the management of relationships between national data holding organisations and local EROs. The alternative is requiring, for example, DWP to deal directly with individual data requests from 380 EROs.

This central organisation could be the eventual system owner for the IER Digital Service, which will manage the 'verification process' – the checking of electors' personal identifiers between EROs and DWP. However, in contrast to verification, data mining is unlikely to be as automated a process.

There should only be further data mining testing on the understanding and acceptance of the need for an ongoing central presence (and any related costs) in order to receive the data from EROs and the national organisation, match it and return the results to each local area.

Any further testing should also be considered in relation to the priority of the overall transition to IER. Plans should therefore take into account the capacity of all the organisations and individuals required to test data mining, specifically to ensure that any testing would not adversely affect their existing commitments to delivering IER.

In addition, there were numerous issues in this pilot with the communication and support provided by Cabinet Office. It is important that **Cabinet Office considers what lessons can be drawn from this pilot, particularly in terms of engagement with EROs, for the wider implementation of IER.**

If further testing is undertaken then, in relation to specific databases:

- **There would be merit in re-testing the Department of Education database** as there were fewer issues than for the other databases and limited results were returned. However:
 - it would be sensible to explore the approach to addressing on this database ahead of any full pilot

- 16 year olds who are under attainer age should be excluded from the data returned to EROs
- **The Welsh Department for Education and Skills database is not suitable for the purposes of data mining and should not be tested again.**
- **The Department of Work and Pensions database should only be included in further testing if:**
 - **full integration of UPRNs is completed** (i.e. all records have UPRNs)
 - **the record currency can be restricted to those with address changes within the past three months**
- **The Royal Mail database should only be included in further testing if:**
 - **the names of individuals below the age for registering to vote are excluded from the data, which Royal Mail has confirmed it can do**
 - **the record currency can be restricted further and the data shared includes the start date of the redirection**
- **The Student Loans Company database should only be included in further testing if:**
 - **the addressing issues experienced in this pilot can be resolved**
 - **testing takes place during October - November or January - February rather than at the end of the academic year**

In relation to any further testing in general:

- **There needs to be a clear understanding of the databases being accessed and a clear data specification provided to the data holding organisations** (based on the requirements of the pilot).
- **For any new database proposed for data mining testing, their approach to addressing should be assessed in advance of the pilot.** There would be limited value at this point in testing a database which lacks UPRNs and has poor addressing information.
- **For any database tested, the potential for returning records to EROs with the original register address attached should be explored** (rather than the address held on the national database). This could, where available, be achieved using UPRNs.
- **Any combination of databases needs to be less complex.** For example, in this pilot data from two databases was combined into one file, and pilot areas could receive more than one file. Many areas could not clearly report on the results and this made evaluating the pilot more

difficult. **For future testing, only one file with data from one database should be provided to each ERO.**

- **There should be mandatory checking of the national data provided against data held locally.** This would allow for an assessment of the added value to EROs of access to national data, as compared to local data which they already have access to. For example, if DfE data is included in the re-testing, it should be compared with locally held education data to assess whether the unregistered individuals identified on DfE could be identified using local data instead.
- **Cabinet Office need to ensure that they maintain good communication between themselves, the data holding organisations and EROs throughout the process,** including after data from the national databases has been returned to EROs.

County data mining

The results from this pilot do not show how useful it would be for EROs for lower tier authorities to have access to data held by an upper tier. However, **EROs in these authorities should be given the legal right of access to data held by upper tier authorities, to put them in position analogous to EROs in unitary authorities.**

EROs are responsible for deciding which local data they are prepared to use in maintaining their register. These decisions should be based on an assessment of the quality of the specific database to be used.

1 Introduction

- 1.1 This report sets out the findings of the Electoral Commission’s evaluation of the 2013 data mining pilot. The pilot, run by Cabinet Office, involved comparing electoral registers with five national public databases and data held by four county councils in order to identify unregistered electors.
- 1.2 The aim of this pilot was to test whether providing Electoral Registration Officers (EROs) with information from these databases could help improve the completeness of their registers.⁶ The pilot targeted groups of electors known to have lower than average levels of registration.

The electoral registers

- 1.3 Electoral registers underpin elections by providing the list of those who are eligible to vote. Those not included on the registers cannot take part in elections. Registers are also used for other important civic purposes, including selecting people to undertake jury service and calculating electorates to inform Parliamentary and local government boundary reviews, which are the basis for ensuring representative democracy.
- 1.4 In addition, credit reference agencies may purchase complete copies of electoral registers, which they use to confirm addresses supplied by applicants for bank accounts, credit cards, personal loans and mortgages.
- 1.5 Great Britain does not have one single electoral register. Rather, each local authority appoints an ERO who has responsibility for compiling an accurate and complete electoral register for their local area.^{7 8}

⁶ By completeness, we mean that ‘every person who is entitled to have an entry in an electoral register is registered’. The completeness of the electoral registers refers to the percentage of eligible people who are registered at their current address. The proportion of eligible people who are not included on the register at their current address constitutes the rate of under-registration.

⁷ In Scotland, in some cases the Assessor of the Valuation Joint Board has been appointed to act as ERO for several neighbouring local authorities.

⁸ By accuracy, we mean that ‘there are no false entries on the register’. The accuracy of the electoral registers is a measure of the percentage of entries on the registers which relate to verified and eligible voters who are resident at that address. Inaccurate register entries may relate to entries which have become redundant (for example, due to people moving home), which are for people who are ineligible and have been included unintentionally, or which are fraudulent.

Current system of updating the electoral registers

- 1.6 At present, EROs use an annual household canvass and rolling registration to update their registers. Electors can register to vote throughout the year (including up to 11 working days before each election) by completing a rolling registration form and submitting it to their ERO. However, most updates to the registers take place during the annual canvass, which is undertaken each autumn. At its simplest, the canvass involves delivering a registration form to each household and following up, via postal reminders and personal visits, those households who do not respond. Revised registers are published by 1 December each year.
- 1.7 The majority of EROs also use locally held data, such as council tax and housing records, to improve the effectiveness of their registration activity. The extent and sophistication of this use varies widely.

Accuracy and completeness of the electoral registers

- 1.8 Previous Electoral Commission research has provided estimates of the accuracy and completeness of the electoral registers. As at April 2011, we estimated the local government registers were 85% accurate and 82% complete.⁹ This equates to approximately 8.5 million unregistered people in Great Britain. However, this does not mean that these registers should have had 8.5 million more entries, because many, but not all, of those not registered correctly may still have been represented on the registers by an inaccurate entry (for example, at a previous address).
- 1.9 Data mining is intended to help improve the completeness of the electoral registers by identifying unregistered electors. The principle behind data mining is that data held by national or local bodies may include people who are eligible to vote but who are not registered. By comparing these databases with the electoral registers, it should be possible to produce a list of potential new electors, and the ERO can then invite them to register.
- 1.10 Data mining may help to improve the accuracy of the registers as well, for example by enabling EROs to identify and remove out of date entries. However, this pilot was not designed to test the effects on accuracy and pilot areas were not asked to identify, or report on, out of date entries.

⁹ The Electoral Commission, *Great Britain's electoral registers 2011* (December 2011). http://www.electoralcommission.org.uk/_data/assets/pdf_file/0007/145366/Great-Britains-electoral-registers-2011.pdf The rates for the April 2011 Parliamentary registers were similar at 86% accurate and 82% complete.

Individual electoral registration

- 1.11 The electoral registration system in Great Britain is changing from household registration to individual electoral registration (IER).¹⁰ At present, one person in every household is responsible for registering everyone else who lives at that address (although individuals can apply to register as well). Under IER, the responsibility for registering to vote will rest with each individual. In addition, in order to be registered each individual will have to provide personal identifiers: their date of birth and National Insurance number. These changes are intended to improve the accuracy of the registers and to modernise the security of the registration and voting systems. The Commission has been calling for IER since 2003.
- 1.12 The previous UK Government, during the passage of the Political Parties and Elections Act 2009 (PPEA), introduced legislation providing for the phased introduction of IER. PPEA included provisions to allow data matching pilots to be carried out, with a view to establishing which national public databases might be useful to EROs in helping to maintain electoral registers during and after the transition to IER.
- 1.13 The Coalition Agreement reached by the Conservative and Liberal Democrat parties set out the Government's plans to speed up the implementation of IER. The Electoral Registration and Administration Act 2013 (ERA) provides the legal framework for IER to be introduced on that basis.

Transition to IER

- 1.14 The final household canvass will take place in spring 2014.¹¹ In summer 2014 all names and addresses on the electoral registers will be compared with records held by the Department for Work and Pensions (DWP) in order to verify the identity of the people on the registers ('confirmation').¹² All electors whose details are matched will be confirmed directly onto the first IER registers and will not have to provide personal identifiers.
- 1.15 All electors whose entries are not matched will be asked to re-register by providing personal identifiers. In addition, from the start of the transition, any new elector will need to make an individual application and provide personal identifiers, as will any electors who change address (even if they have been confirmed at their previous address).

¹⁰ Northern Ireland has used a system of individual electoral registration since 2002.

¹¹ There will still be annual canvasses under IER, but these will take a different form.

¹² The timing for the introduction of IER in Scotland is different as a result of the referendum on independence in September 2014. The confirmation process will take place in late September with a subsequent write out to unconfirmed electors beginning in October 2014.

1.16 Under current plans, electors on the 2014 registers who are not confirmed will have until December 2016 to provide personal identifiers, before they are removed.¹³ However ERA enables Ministers to lay an Order before the UK Parliament to provide for the transition to be completed by December 2015 instead. The Government has made it clear that its intent is to complete the transition in 2015. Therefore, while there is uncertainty as to whether the point of removal of electors that have not provided personal identifiers will be in 2015 or 2016, it is our view that EROs should plan on the basis that they will have to be ready for the point of removal to be 2015.

Previous pilot schemes

2011 data matching pilot¹⁴

1.17 The first pilot held under PPEA was conducted in 2011. It involved matching electoral registers from 22 pilot areas against ten national databases in order to test whether giving EROs access to the data would help them improve the accuracy and completeness of their register.

1.18 Under PPEA, the Commission has a statutory responsibility to report on the effectiveness of data matching pilots. Our report into the 2011 pilot was critical of the methodology of the pilot. Issues included the complexity of the data returned to EROs; the lack of a consistent definition of what constituted a match between the registers and the databases; and the overlap between the timing of pilot and the annual canvass.¹⁵

1.19 The methodological flaws meant that it was not possible for us to conclude whether access to national databases could assist EROs in maintaining their electoral registers. However, we said that there was merit in re-testing nearly all of the databases used in the pilot but in a way which allowed for the collation of meaningful results.

¹³ Any elector with an absent vote (postal or proxy voters) will need to be confirmed or provide their personal identifiers before the revised electoral registers are published by 1 December 2014 in order to retain their absent vote. However, they will still be able to vote in a polling station until either December 2015 or December 2016, depending on the end date of the transition.

¹⁴ The terminology used for these pilots has changed over time. Initially, the 2011 pilot, which aimed to identify unregistered electors, was referred to as a data matching pilot. Subsequently, data matching was adopted as an umbrella term covering two different processes: 1. Confirmation – matching register entries to DWP data to confirm identity and retain electors on the registers. 2. Data mining – using data to find unregistered electors.

¹⁵ The Electoral Commission, *Data matching schemes to improve accuracy and completeness of the electoral registers – evaluation report* (March 2012) http://www.electoralcommission.org.uk/_data/assets/pdf_file/0010/146836/Data-matching-pilot-evaluation.pdf

1.20 The recommendations from our evaluation of the 2011 pilot are referred to throughout this report where relevant.

2012 confirmation pilot

1.21 The results from the 2011 pilot did indicate that a majority of existing electors could be matched against the DWP database. This finding was the basis for the principle of confirmation and was the subject of a separate pilot in 2012.

1.22 The confirmation pilot involved matching the electoral registers from 14 pilot areas against data held by DWP. On average, over 70% of electors could be matched. Our evaluation concluded that confirmation was an effective and reliable way of verifying the identities of electors and should be used during the transition to IER. However the results also showed that confirmation was less effective for certain groups of electors, particularly students and those who move home frequently.¹⁶

2013 pilot: aims and objectives

1.23 The role of data mining during or after the transition to IER is not yet clear. The rationale behind the pilot schemes is that it potentially offers a way for EROs to target registration activities and so make more effective use of resources.

1.24 The current pilot was designed to test whether providing EROs with access to this data would enable them to identify, and register, currently unregistered individuals. The pilot focused on three groups of electors known to have lower than average rates of registration: attainers, home movers and students (see Chapter 2 for further information on why these groups were selected).

1.25 The pilot was also intended to help assess the cost-effectiveness of data mining.

The Electoral Registration Data Schemes (No. 2) Order 2012

1.26 The Electoral Registration Data Schemes (No. 2) Order 2012 ('the 2012 Order'), made on 19 December 2012, allowed for the sharing of specified data between the named pilot areas and data holding organisations. The data holding organisations named in the Order are:

¹⁶ The Electoral Commission, *Data matching pilot – confirmation process* (April 2013) http://www.electoralcommission.org.uk/_data/assets/pdf_file/0009/154971/Data-matching-schemes-confirmation-process-evaluation-report.pdf

- the Department for Work and Pensions
- the Department for Education
- the Welsh Department for Education and Skills
- the Student Loans Company Ltd
- the Higher Education Funding Company for England¹⁷
- Royal Mail Ltd
- Cumbria, Hampshire, Nottinghamshire and Lancashire county councils

1.27 The Order also specified which pilot areas each data holding organisation could share data with.

1.28 Under the 2012 Order, an agreement between the relevant data holding organisation and pilot area needed to be in place before personal data could be shared between the two parties. The purpose of the agreement was to set out: governance arrangements for data transfer and matching; the expected inputs and outputs; information security standards; and timescales.

This evaluation

1.29 The Commission has a statutory responsibility to report on the effectiveness of the data matching schemes. Sections 35 and 36 PPEA outline the aspects that should be covered in our evaluation, including whether the scheme could assist EROs in meeting their registration objectives, the administration and cost of the scheme and any public objections. Based on this, we have developed specific criteria with which to evaluate the data mining pilot.

1.30 Our evaluation aims to assess:

- whether the national and/or county council databases could provide EROs with current information on eligible and unregistered electors within each of the target groups
- the number of new registrations achieved from the individuals identified by these databases
- whether there were any public objections to the scheme

¹⁷ Although the Higher Education Funding Company for England was named in the Order, they did not participate in the pilot.

- the cost of using data mining to identify eligible but unregistered electors

Sources for our evaluation

1.31 Our evaluation draws on:

- interviews with all of the pilot areas, all of the national data holding organisations and one county council
- results submitted by the pilot areas
- information provided by the pilot team in the Cabinet Office

1.32 The Commission would like to thank everyone who has participated in the pilot and helped with this evaluation, particularly the staff at the local authorities and Valuation Joint Boards.

This report

1.33 The rest of this report sets out further details of how the pilot worked and presents the results, analysis and associated recommendations. The chapters are arranged as follows:

- Chapter 2 looks at the set up of the pilot including the databases accessed
- Chapter 3 sets out the processes used in the pilot including data transfer, matching and follow up
- Chapter 4 provides information on costs associated with the pilot
- Chapter 5 sets out our analysis of the national data mining activities
- Chapter 6 sets out our analysis of the county council data mining activities
- Chapter 7 sets out our recommendations
- Appendices A to F set out our detailed findings in relation to the databases tested in this pilot

2 Set up – target groups and databases

2.1 This chapter explains the set up and matching process for the national and county data mining.

Key points

- The pilot involved targeting three groups of electors known to have low registration rates – attainers (17 year olds and some 16 year olds), home movers and students.
- Eighteen EROs were involved in the national pilot. They were provided with data held on five national public databases. Each of the EROs accessed data from at least two of these databases.
- In addition, four district council EROs were given access to data held by their respective county council, with a specific focus on trying to identify unregistered attainers.

National data mining

Target groups of electors

2.2 Three groups of electors were targeted in this pilot: attainers, home movers and students. These are all groups which are known to have lower than average rates of registration.

2.3 There are several likely reasons for these lower registration rates such as high population mobility or disaffection with traditional politics. It is possible that data mining could help to register those people who are not on the register because they move house frequently but it is likely to be less effective at registering those who feel disengaged from traditional politics.

Attainers

2.4 An elector who is not yet 18 years of age must be shown on the register with the date on which they will attain the age of 18. Those electors are

called attainers as they are about to attain voting age. Attainers are predominantly 17 year olds, with some 16 year olds.¹⁸

- 2.5 Our research has found that registration levels varied by age, with 55% of 17-18 year olds registered at their current address, compared to 86% for electors aged 35-54 and above 90% for older age groups.¹⁹

Home movers

- 2.6 Our research found that the registration rate for electors who had been at their current address for up to one year was 26%. This is substantially lower than for electors who had been at their address for between one and two years, where the registration rate was 76%.

Students

- 2.7 It is difficult to produce robust estimates of student registration rates, since many students are entitled to be registered twice (at both their home and term-time addresses). The Commission's 2005 research into electoral registration, using 2001 census data, estimated the student registration rate to be around 78%.²⁰
- 2.8 However, registration rates were lower at term-time addresses and substantially lower for students who had moved in the previous six months (around 55%).

The databases

- 2.9 The databases used in this pilot were selected in response to the target groups identified, e.g. education data was considered a useful source of data on attainers. However, the final list of databases involved in the pilot was also determined by whether data holding organisations wanted to participate. For example, the Driver and Vehicle Licensing Agency was involved in the 2011 pilot but declined to participate this time around. In addition, there was an initial ambition to include some NHS data but no agreement was reached on access.

¹⁸ An attainer is an individual who will turn 18 before the end of a 12 month period starting from the next 1 December after the application is made i.e. if an application is made in spring 2013, the applicant will be eligible as an attainer if they turn 18 any time before December 2014.

¹⁹ The Electoral Commission, *Great Britain's electoral registers 2011* (December 2011). http://www.electoralcommission.org.uk/_data/assets/pdf_file/0007/145366/Great-Britains-electoral-registers-2011.pdf

²⁰ The Electoral Commission, *Understanding Electoral Registration* (September 2005) http://www.electoralcommission.org.uk/_data/assets/pdf_file/0020/47252/Undreg-FINAL_18366-13545_ENSW.pdf

2.10 Five national public databases were used in this pilot and these are shown in the table below against the target groups they were intended to identify.²¹

Table 1: National databases and target groups

| Target group | Databases |
|--------------|--|
| Attainers | Department for Work and Pensions (DWP) Department for Education (DfE) Welsh Government, Department for Education and Skills (DfES) ²² |
| Students | Student Loans Company (SLC) DfE / DfES |
| Home movers | DWP Royal Mail |

2.11 Table 2 below summarises the coverage and sources of the national databases used in the pilot. It also indicates what, if any, currency restrictions were placed on the data used in this pilot. One of the issues identified in the 2011 pilot was that much of the information provided from the national databases was not current enough for the purposes of electoral registration. The currency restrictions employed in this pilot were an attempt to reduce this problem.

²¹ Data held by the Higher Education Funding Company for England (HEFCE) was also intended to be used in this pilot. However, Cabinet Office and HEFCE could not reach agreement on the terms and conditions for sharing data, as required by the 2012 Order (see paragraph 1.28), within the timescales of the pilot.

²² Cabinet Office had planned to use two datasets held by the Welsh Department for Education and Skills (DfES): the School Census described in Table 2 below and also the Lifelong Learning dataset. DfES provided an extract of this second dataset to Cabinet Office, however there were a number of formatting issues which could not be resolved in time for it to be used in this pilot.

Table 2: National databases used in the 2013 data mining pilot

| Data-holder | Database | Coverage of data used in pilot | Currency restrictions | Sources and updates |
|---|--|--|---|--|
| Department for Education | National Pupil Database | Schools, academies and some non-mainstream educational provision in England. Does not cover independent schools or further education colleges. | Records from the October 2012 census only | Mostly submitted directly by schools although in some areas comes through the local authority. The census is conducted three times a year. |
| Welsh Department for Education and Skills | Pupil Level Annual School Census | All local authority maintained schools in Wales (covering approx. 97% of pupils) | Records from the January 2013 census only | Submitted by schools via local authorities. The census is conducted once a year. |
| Department for Work and Pensions | Customer Information Systems (social security, tax credits and child benefit records only) | All individuals with a National Insurance number or a child reference number | Address updated within last 12 months | Various databases in DWP and HM Revenue and Customs. Updated through customer or employer contact. |

| Data-holder | Database | Coverage of data used in pilot | Currency restrictions | Sources and updates |
|---------------------------|--|---|---|--|
| Student Loans Company Ltd | Student Finance Customer Account System (Higher Education) | Students undertaking undergraduate higher education qualifications, who have applied for a student loan ²³ | Current students only | Student initiated. Students make a fresh finance application each academic year. SLC conducts identity checks before approving initial application. Attendance at the specified institution is checked three times a year. |
| Royal Mail Ltd | Mail redirection service (Change of Address Update and Suppress databases; Home Mover Mailing Service) | National but optional – individuals apply for mail to be redirected when they move home. | Redirections set up within the last 2 years | Individuals submit request to have mail redirected for a set period of time, providing previous and new addresses. Royal Mail conducts identity checks before approving application. |

²³ There will also be a small number of postgraduate students who are eligible to apply for student loans e.g. on courses such as teacher training.

The pilot areas

- 2.12 A total of 20 areas participated in the pilot: 13 local authorities in England, five local authorities in Wales and two Valuation Joint Boards in Scotland. Of these, 18 were involved in the national data mining; two of the authorities in England looked only at county council data.
- 2.13 All areas who participated in the 2011 or 2012 pilots were invited to participate in this pilot. In total, 13 of this year's pilot areas were involved with one or both of the previous pilots. The remaining seven areas had previously expressed interest in participating in a data matching pilot.
- 2.14 The pilot areas were not selected to be representative of Great Britain but do offer a reasonable spread of different types and sizes of authority across England, Scotland and Wales
- 2.15 For the national data mining, pilot areas looked at different groups of electors. As can be seen from Table 3 below, there was not an even distribution between the different groups. Nine pilot areas tested the databases for mining attainers, five tested the student databases and 14 the databases for home movers.
- 2.16 This means that some pilot areas tested the data for more than one group; in fact, some areas received as many as three different sets of data.

Table 3: National data mining – pilot areas and target groups

| Area ²⁴ | | Attainers | Home movers | Students |
|--------------------|--|-----------|-------------|----------|
| England | Coventry | ✓ | | ✓ |
| | Greenwich | | ✓ | ✓ |
| | Harrow | ✓ | | |
| | Richmond-upon-Thames | | ✓ | |
| | Rushmoor | ✓ | | |
| | South Ribble | | ✓ | |
| | Southwark | | ✓ | |
| | Sunderland | ✓ | ✓ | |
| | Tower Hamlets | ✓ | ✓ | ✓ |
| | Wigan | | ✓ | |
| | Wolverhampton | ✓ | | ✓ |
| Wales | Ceredigion | ✓ | ✓ | ✓ |
| | Conwy | | ✓ | |
| | Pembrokeshire | ✓ | ✓ | |
| | Powys | ✓ | ✓ | |
| | Wrexham | | ✓ | |
| Scotland | East Renfrewshire / Renfrewshire VJB ²⁵ | | ✓ | |
| | Lothian VJB | | ✓ | |

²⁴ In addition, Barrow-in-Furness and Mansfield took part in the pilot looking solely at county data.

²⁵ Renfrewshire Valuation Joint Board participated in the pilot using only the register for East Renfrewshire (the VJB covers East Renfrewshire, Inverclyde and Renfrewshire).

County data mining

- 2.17 This part of the pilot involved four district councils receiving data held by their respective county council.
- 2.18 In many parts of England, local government is organised as a multi-tiered structure. Each county will have a single county council and several district councils within the county area.
- 2.19 Local government functions are shared between the county and district tiers. For some functions, both tiers will have responsibility whereas for other functions responsibility will lie with only one tier. For example, district councils are responsible for electoral registration and county councils are responsible for education.
- 2.20 Unlike in a unitary council (as in all of Wales and many areas in England, where a single authority has responsibility for all local government functions) an ERO for a lower tier authority has no legal right to access the data held by the county council.
- 2.21 The main aim of this part of the pilot was to see if it was possible to establish data sharing arrangements between a county and district council for the purposes of electoral registration.

Areas, target groups and databases

- 2.22 Four pilot areas participated in the county data mining and all focused on identifying attainers. Three of the county councils supplied their education database, with one providing their broader children’s services database.
- 2.23 The coverage of each of the county council datasets is summarised in Table 4 below.

Table 4: County data mining – pilot areas and databases

| Pilot area | County council | Dataset |
|-------------------|-----------------|--|
| Barrow-in-Furness | Cumbria | Education data |
| Mansfield | Nottinghamshire | Children’s services data |
| Rushmoor | Hampshire | Education data |
| South Ribble | Lancashire | Education data – records from the October 2012 school census |

Conclusion

- 2.24 Although less complex than the 2011 pilot, this pilot was still ambitious in terms of the number of different databases it was trying to assess. For the national data mining, there were two databases for each target group of electors. Half of the pilot areas were looking at more than one target group, with two looking at all three of attainers, home movers and students.
- 2.25 This made some of the practical arrangements more complex, for example by increasing the number of legal agreements required (a separate agreement was needed for each pairing of data holding organisation and pilot area).

3 Pilot processes – data transfer, matching, follow up

3.1 This chapter explains the overall processes followed in the pilot: data transfer, data matching and the follow up work conducted by pilot areas. We also set out some of the associated issues for our evaluation.

Key points

- There were considerable delays to the original timetable for establishing this pilot. A significant cause of the delays was the lack of capacity within Cabinet Office due to their workload related to the transition to individual electoral registration (IER).
- The delays affected the amount of work the pilot areas could do with the data.
- Three different organisations were involved in matching the national databases against the electoral registers in order to produce a list of potential unregistered electors. Each organisation used similar but not identical matching criteria. We have not been able to quantify any variation in the matching caused by differences in the matching criteria.
- For the national data mining, Cabinet Office's original intention was that pilot areas should adopt a fairly standardised approach to checking the data received and contacting the individuals identified, to ensure that results were comparable. In practice, however, the nature and extent of follow up work varied widely.
- Much of this variation was caused by practical difficulties, for example the need to spend more time than expected in ensuring the accuracy of the data received. However, some of the variation could have been avoided if there had been fewer delays and a greater level of support provided by Cabinet Office to pilot areas. In particular, a few areas told us they felt unsupported and were unclear about what to do.
- Although the wide variation in follow up activity has limited our evaluation in certain ways, we do have more complete and comparable data than in the 2011 pilot and so are able to reach more definite conclusions.

Introduction

- 3.2 The Cabinet Office was responsible for the management of the pilot. They coordinated the legal agreements and transfer of data and provided support to the pilot areas.

Timetable and delays

- 3.3 The original timetable for the pilot involved pilot areas sending their registers to the Government Digital Service (GDS) by early February 2013. The matching would be conducted in February, allowing GDS to compile and return the results in early March, well in advance of 17 April, the statutory deadline for data transfer.²⁶ Pilot areas would then submit results from their follow up work by the end of April, several months before our statutory evaluation deadline of 17 July.
- 3.4 In practice, the registers were sent in early March, the matching was conducted in late March and the results returned only a day or two before the 17 April cut-off point.
- 3.5 The delays were in part caused by difficulties in finalising legal agreements between the data holding organisations and pilot areas (see paragraph 1.28). However, it seems that these difficulties were largely restricted to the agreements with Royal Mail and the county councils. Most of the pilot areas and data holding organisations we spoke to indicated that the process was fairly straightforward, with many simply updating the agreements used in previous pilots.
- 3.6 However, the pilot also suffered from the workload created by Cabinet Office's wider Electoral Registration Transformation Programme which is responsible for introducing IER, in particular the demands on GDS around the development of the IER Digital Service.²⁷
- 3.7 If there is any further work on data mining, we strongly recommend that Cabinet Office creates realistic project plans and timescales, making sure in advance that they have sufficient resources and staff capacity to meet the stated timescales. Any planning should consider the requirements imposed by the overall work required as part of the implementation of IER as that will always take priority over pilots.

²⁶ The SI which enabled the pilot set 17 April as the deadline for data transfer.

²⁷ The IER Digital Service is the central IT system being developed as part of the transition to IER. It will be responsible for managing the confirmation process as well as the transfer of data between Electoral Registration Officers (EROs) and DWP as part of the ongoing verification of electors' personal identifiers.

Impact of the delays

3.8 The impacts of the delays include:

- GDS having little time to investigate errors or clean returned data.
- GDS not having time to restrict results to the postcodes or wards specified by some of the pilot areas (so contributing to the higher volume of records returned).
- Results being returned during the election period, when five pilot areas were running elections.²⁸ These areas were therefore unable to do any work with the data until the elections were complete.²⁹
- Generally compressing the time available for follow up work with potential electors.
- Pilot areas having little time at the end of the process to provide results or to fix errors in the data reported to us.

3.9 The rush to transfer data before the deadline of 17 April meant that one pilot area (Pembrokeshire) did not receive the data they expected for attainers. Another area (South Ribble) did not do any follow up work during the timescales of the pilot so we have no results to report for them.

Data transfer

3.10 Data was transferred between the various organisations by a combination of secure courier and secure email.

3.11 Both of the previous data matching pilots used secure courier to transfer data. For any national roll out of data mining this would not be a sustainable or cost effective way of transferring the data due to the volumes of data involved.

3.12 It is a positive step that secure email was used for some of the data transfer in this pilot, although it was limited to government bodies (including local authorities) with a pre-existing secure government email account. Some of the pilot areas who tried to set up a secure account as part of this pilot reported that it was not a particularly quick process, and

²⁸ Barrow-in-Furness, Mansfield, Rushmoor and South Ribble had county-wide elections. Wolverhampton had a local by-election.

²⁹ The priority in those areas at that time was running the elections. They also could not register people from 11 working days before polling day and did not want to write out to people and potentially have them believe they would be registered for the elections when they would not have been.

not all completed it in time for use in the pilot. It is possible that the work being carried out to establish connectivity between all local authorities and the IER Digital Service means this would be less of an issue in the future.

- 3.13 The non-governmental data holding organisations (Student Loans Company and Royal Mail) had to transfer their data by secure courier, which they found cumbersome. Both would have preferred to use secure electronic transfer, which is typically used in other data sharing arrangements they are part of.

National data mining

Matching process

- 3.14 Each database was matched against the relevant electoral registers in order to identify the individuals who were already registered.³⁰ This left a list of individuals who were on the database but could not be found on the electoral registers – at least, not at that address. This process took place centrally, although with several organisations involved.

Responsibilities

- 3.15 The pilot areas first sent an extract of their electoral registers to the GDS, who carried out some basic standardisation.³¹
- 3.16 The matching was conducted by GDS, the Department for Work and Pensions (DWP) and Transactis (on behalf of Royal Mail).³²
- GDS matched the electoral registers against data held by the Department for Education (DfE), the Welsh Government Department for Education and Skills (DfES) and the Student Loans Company (SLC).
 - DWP matched the electoral registers against their Customer Information Systems database.
 - Transactis matched the electoral registers against the Royal Mail databases.
- 3.17 The results from all databases were then compiled by GDS. All of the results for each group of electors were compiled into one file per pilot

³⁰ An extract of each database was used covering the relevant geographical area.

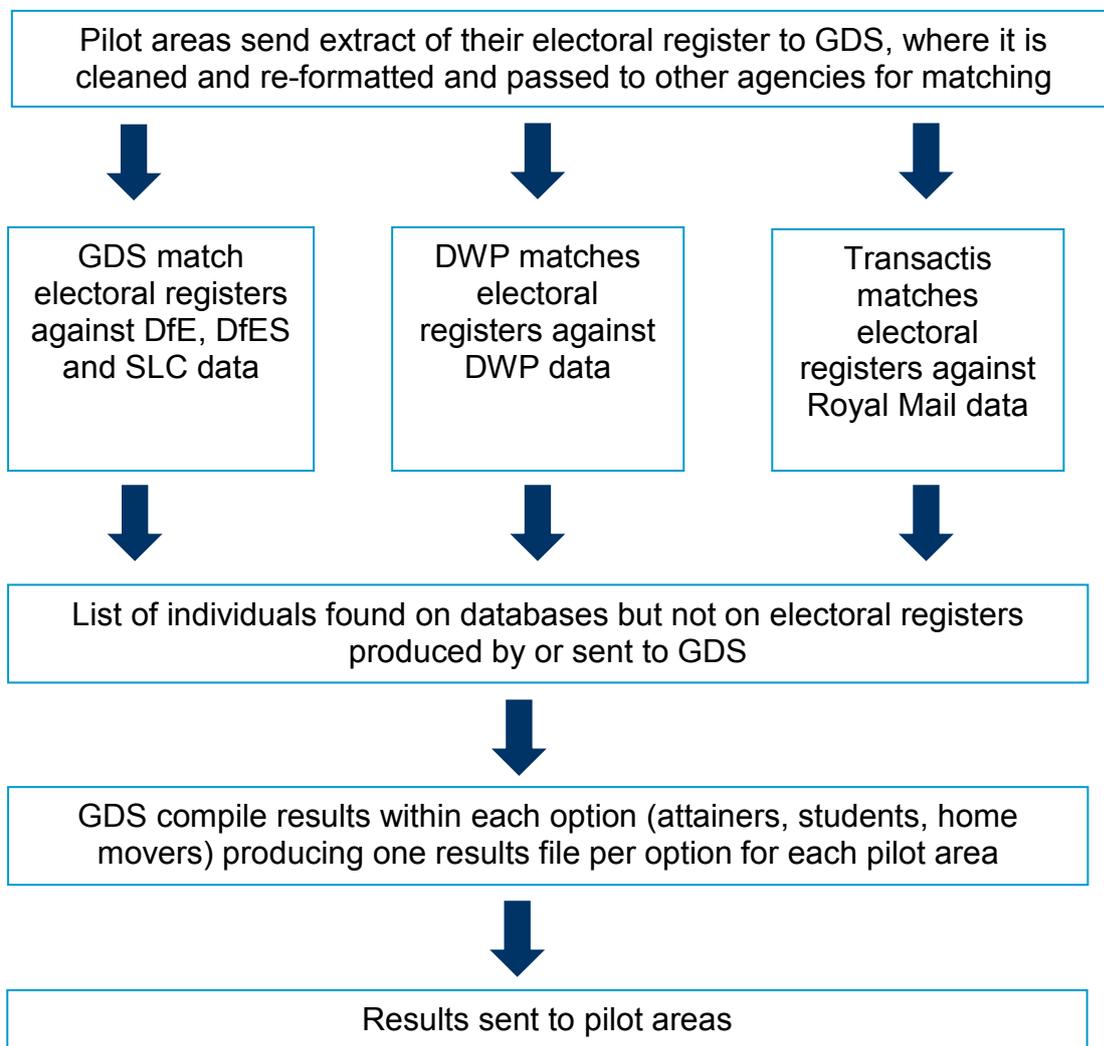
³¹ Ensuring that e.g. field headings are identical, key bits of data are in the same field for each register.

³² CDMS Ltd t/a Transactis are a pre-selected supplier under HM Government's Data Access, Processing and Analytics Framework. Transactis are Royal Mail's preferred data management partner and are contracted to do other matching work as well as this pilot.

area. For example, for home movers, GDS compiled the results from DWP and Transactis into a single file for each of the 14 pilot areas. This stage also involved identifying individuals who had been returned from more than one database and removing duplicates. The data returned to pilot areas indicated where an individual had been identified on more than one database.

3.18 The different groups (attainers, home movers and students) were treated separately at this stage. Pilot areas who selected more than one group therefore received more than one results file. There was no checking for duplicates between different target groups for the same pilot area. For example, there was no mechanism to identify and remove duplicate names between the records returned for home movers and students.

Figure 1: National data mining – matching process



Matching algorithm – DWP and GDS

3.19 DWP and GDS used automated matching processes to conduct the matching although each separately developed a matching algorithm (the rules that determine who matches).

3.20 DWP used the algorithm that they developed as part of the 2012 confirmation pilot while GDS developed their own algorithm based on the same principles.

3.21 This algorithm involves a two-stage match process: address then name. The algorithm first attempts to match the address, using either the UPRN or address lines (e.g. 1 Acacia Avenue).³³ If the address could be matched, the algorithm then attempts to match the names of the individuals living at that property, using different combinations of first name, middle name and surname.

3.22 The results are classified into red, amber or green matches.³⁴

- Red: no address match, or an address match but no name match.
- Amber: address match and partial (or ‘fuzzy’) name match, for example exact first name, fuzzy last name (for example, where there was a slight difference in spelling between the two data sets).
- Green: address match and exact name match, for example full first name and full surname.

3.23 The purpose of data mining is to identify individuals who are not on the registers, i.e. the individuals who did not match during this process. The results sent back to the pilot areas were therefore a list of the red matches from the relevant databases – the records on that database that could not be matched against an entry on the registers. This is a significant improvement on the 2011 pilot, where pilot areas were provided with all the names on the national database and left to determine their own definitions of what constituted a match, leading to inconsistency between areas.

³³ UPRNs are unique 12-digit codes assigned to each property at a local level. These local lists are then combined to form the National Land and Property Gazetteer in England and Wales and the One Scotland Gazetteer in Scotland. Out of the databases used in this pilot, DWP is the only one which holds UPRNs, with approximately 88% of the records having UPRNs. The matching for the four other databases therefore compared the address fields held on that database with those held on the relevant electoral register.

³⁴ Our report on the 2012 confirmation pilot contains further details of the matching algorithm and classification. The Electoral Commission, *Data matching pilot – confirmation process* (April 2013) http://www.electoralcommission.org.uk/data/assets/pdf_file/0009/154971/Data-matching-schemes-confirmation-process-evaluation-report.pdf

- 3.24 Amber matches were not included in the results sent back to pilot areas as evidence from the confirmation pilot suggested that a large proportion of these records were existing electors but had failed to match due to minor differences of address or name.

Matching algorithm – Transactis

- 3.25 Transactis also used an automated matching process to match the electoral registers against the Royal Mail data, using the same algorithm that they use for all of their Royal Mail data matching work.
- 3.26 This algorithm is a single stage process which involves matching data held in all fields at once (i.e. matching name and address at the same time). The algorithm has provision for fuzzy matching as well as rules to assist matching more complex addresses, such as flats and tenements.
- 3.27 Transactis re-formatted the electoral registers that they received to assist the matching process, quite extensively in some cases.

Pilot areas' follow up work

- 3.28 Once each pilot area received their results for the target groups they were focusing on they could then assess the data and carry out follow up work using the names provided. Broadly, this meant either sending letters with a registration application form or door to door canvassing.
- 3.29 One of the main problems with the 2011 pilot was the wide variation in follow up work between the different pilot areas, coupled with the overlap between the pilot and the annual canvass. This meant that results were not comparable and we could not draw any conclusions about the usefulness of the databases.
- 3.30 This year's pilot took place outside of the canvass period which makes interpreting the results more straightforward. In addition, the need for standardised follow up was accepted by Cabinet Office and included in their plans, at least for the national data mining. Cabinet Office intended that each pilot area adopt a similar approach to the follow up work and provided guidance outlining each stage.
- 3.31 In practice, however, there was a considerable amount of variation in the follow up work, although not to the same extent as in the 2011 pilot. Below, we set out the implications of some of that variation for this evaluation. Appendix G contains further details of each pilot area's approach to follow up work.

Sampling

- 3.32 Cabinet Office thought that pilot areas would either follow up all of the records received or work with a random sample if the volume was too high. For the home movers option, pilot areas had been able to select particular wards or postcodes, so they would only get records back for individuals who lived in those locations. In fact, Royal Mail were unable to limit home movers data to the wards or postcodes selected by many

of the pilot areas. The volume of data provided was therefore much greater than the pilot areas expected.

3.33 In practice, many areas took different approaches to identifying which records to follow up. As a result, where we see variation in the numbers of new registrations achieved by different pilot areas, we cannot be sure how much of the variation is due to different approaches to sampling. These differences include:

- **Using a random sample or selecting particular records.**
- **Taking different approaches to identifying and excluding out of date records or records relating to existing electors.** One pilot area did no checks before sending letters, some did very extensive checks, while others conducted limited checks or checked certain records only, for example only those returned from one database.
- **Selecting attainer samples based on varying date of birth ranges.**³⁵ For example, some areas excluded all 18 year olds while others wrote to them anyway, and different areas applied different lower date of birth thresholds.³⁶

3.34 In general, we have more complete results for the areas which conducted more thorough checks before sending letters. This is because they were able to identify more of the individuals who were existing electors, ineligible individuals or no longer resident, than the areas who relied on these individuals (or the new resident at that address) to respond to the letter. Also, response rates to follow up were higher in some of the areas who completed more checks, perhaps because they were writing out to a more refined list.

Method and timing of follow up work

3.35 Pilot areas were able to contact potential new electors by letter or by door to door canvassing. Mostly, the pilot areas wrote to the individuals identified. The Cabinet Office provided a template cover letter (which the Commission had an opportunity to comment on), briefly explaining the purpose of the pilot. This was sent out along with a registration form.

3.36 However, the times at which letters were sent out varied. Some were sent out close to the reporting date and so there may well be additional

³⁵ Cabinet Office specified an age range rather than a more precise date of birth range for the attainer files, meaning that individuals who were already 18 and individuals too young to qualify as attainers were included on this data.

³⁶ It has become apparent that different areas use different cut off points for the lower age bracket for inclusion on the register, including December 1995, February 1996 and January 1997. This issue applies to the pilot areas looking at the county council data as well. It would be helpful for any future pilot to impose a consistent interpretation.

responses which are not included in the figures in this report. In addition, a few areas issued reminders while others did not.

- 3.37 This means that the numbers of new registrations achieved across the pilot areas vary, partly due to the different practices used in follow up, and, again, this makes it more difficult to compare the results in a meaningful way.
- 3.38 Pilot areas were also encouraged to canvass at least some of the non-responders following the write-out stage as we know that door to door canvassing can have a positive impact on response rates.³⁷
- 3.39 However, only four out of the 18 areas did any sort of canvassing, and in all of these it was limited in some way, e.g. to a particular postcode or to individuals identified on one database. This lack of canvassing was primarily a result of the delays in the pilot set up (see paragraphs 3.3 to 3.9).

County data mining

Matching process

- 3.40 This element of the pilot differed from the national data mining in a number of key ways, namely that each pilot area received data from a different database and they conducted the matching process themselves.
- 3.41 It had been envisaged that the pilot areas might develop their own algorithms, similar to those used in the national data mining, either themselves or through their Electoral Management Software (EMS) provider. In the end, only one pilot area worked with their EMS provider to conduct some automated matching. Other areas manually checked the data against the register, in one case using functionality within Microsoft Excel to assist them.
- 3.42 These differences are not a significant problem for these four pilot areas as the aim of this element of the pilot was primarily to explore what some of the potential challenges and benefits would be for lower tier councils in accessing upper tier data.

Follow up

- 3.43 Cabinet Office did not originally specify that these pilot areas should do any follow up work beyond the matching (because the aim of the county

³⁷ The Electoral Commission, *Great Britain's electoral registers 2011* (December 2011) http://www.electoralcommission.org.uk/_data/assets/pdf_file/0007/145366/Great-Britains-electoral-registers-2011.pdf

pilot was different, focusing on whether it was feasible to share this data and whether pilot areas would be able to successfully conduct the matching). In the end, Cabinet Office did encourage pilot areas to contact the individuals they could not find on their register, but this was not required.

- 3.44 In practice there was therefore significant variation in the approach to follow up and reporting.
- 3.45 Three of the pilot areas wrote out to unregistered individuals but, due to the elections, letters were only sent in the middle of May (see paragraph 3.8). One pilot area (South Ribble) did not send any letters out during the pilot timescales, although they are still planning to contact the unregistered individuals. This area has therefore not reported any results.

Support and guidance provided by Cabinet Office

- 3.46 Pilot areas had mixed views on the support and guidance provided by Cabinet Office. While some were positive about the support provided, a few areas said they felt unsupported, were unclear what to do and were left feeling disillusioned with the pilot process.
- 3.47 Several areas said that guidance and communication from Cabinet Office could be improved. Explanations of how to interpret the data and updates on progress (during the delays) were considered to be particularly lacking. For example, some areas told us that they received two versions of the same data file with no clear explanation as to why (one was an updated version with some errors corrected).
- 3.48 Others said there was inadequate explanation of some detailed elements such as the meaning of column headings in spreadsheets. The impact on follow up work includes pilot areas having to spend longer than anticipated making sense of the data before they could start work on it.
- 3.49 Misunderstandings about which database the record had been returned from contributed to not being able to report by database, and one pilot area worked with records returned from only one database due to incorrect data labels and poor communication from Cabinet Office.
- 3.50 This was a pilot and some mistakes could be expected. It is, however, important that Cabinet Office considers what lessons can be drawn from this pilot, in terms of engagement with EROs, for the wider implementation of IER.
- 3.51 In addition, it would have been beneficial if Cabinet Office had provided feedback to the data holding organisations during the course of the pilot,

particularly about any issues that had been identified with their data (see Chapter 5 and Appendices A-F for details of issues with individual datasets).

Conclusions

- 3.52 This pilot suffered from fewer process issues than the 2011 pilot. However, there were still a variety of issues which either directly impacted on the running of the pilot (and therefore its likelihood of success) or on our ability to evaluate it.
- 3.53 Most significantly, the delays at the start of the pilot process meant that there were tight deadlines for data transfer with no contingency period in which to deal with problems. The delays also had a direct impact on the evaluation as pilot areas had less time to conduct and report on their follow up work with potential new electors.

4 Costs

4.1 This chapter sets out our summary of the costs incurred in the pilot.

Key points

- It is not possible to produce an overall figure for the cost of this pilot. This is because we do not have final costs for all pilot areas or any costs for Cabinet Office (including the Government Digital Service), who conducted a lot of the work on the pilot.
- Since we are unable to produce a cost for this pilot, we are also unable to estimate the cost per new elector registered or the likely cost of any national rollout. Any estimates of these would need to include the cost of coordinating and managing the pilot (the role taken by Cabinet Office in this pilot), as any future work with data mining would require some form of central coordination.

Pilot areas

- 4.2 The costs incurred by the areas participating in the pilot are set out in Table 5. The costs for Harrow, Pembrokeshire, Powys and Renfrewshire VJB are estimates as these areas had not submitted final costs in time for inclusion in this report.
- 4.3 As can be seen from the table, there is a wide range in the pilot areas' costs, from a couple of thousand pounds up to £27,000. This variation may be due to the demographics of the area, how many records the pilot area received or the approach to follow up they adopted.

Table 5: Pilot areas' costs

| Pilot area | Total costs |
|-----------------------------|----------------------------|
| Barrow-in-Furness | £16,609 |
| Ceredigion | £33,007 |
| Conwy | £10,965 |
| Coventry | £16,130 |
| Greenwich | £6,680 |
| Harrow | £15,000 |
| Lothian VJB | £7,140 |
| Mansfield | £2,654 |
| Pembrokeshire ³⁸ | £9,000 |
| Powys | £27,000 |
| Renfrewshire VJB | £8,300 |
| Richmond-upon-Thames | £10,971 |
| Rushmoor | £8,419 |
| South Ribble | No estimate or final costs |
| Southwark | £7,036 |
| Sunderland | No estimate or final costs |
| Tower Hamlets | £27,157 |
| Wigan | £11,394 |
| Wolverhampton | £10,990 |
| Wrexham | £11,513 |
| Total | £239,965 |

³⁸ Pembrokeshire also submitted an estimate of £6,500 for work on the attainments option, but they did not receive this file and so we have not included this figure in their total costs.

Data holding organisations and GDS

4.4 Only Royal Mail charged Cabinet Office for accessing their data (plus charges for data processing and matching), with the Department for Work and Pensions charging for data processing and project management. The other three national data holding organisations and the four county councils did not charge. However there was a cost associated with the unsuccessful negotiations to access the Higher Education Funding Council for England's data (see paragraph 2.10) as well as the cost of using secure couriers to transfer data between some of the data holding organisations and pilot areas.

Table 6: Data holding organisations' costs

| Data holding organisation / item | Cost |
|--|----------------|
| Department for Education | - |
| Welsh Department for Education and Skills | - |
| Department for Work and Pensions | £28,992 |
| Higher Education Funding Council for England | £1,204 |
| Royal Mail | £40,869 |
| Student Loans Company | - |
| Secure couriers | £12,887 |
| Total | £83,952 |

4.5 We asked Cabinet Office for a summary of their costs – to cover the time spent managing the pilot and also the work conducted by GDS (who are part of Cabinet Office), who did a substantial amount of the matching and data-processing (see paragraphs 3.14-3.18). However, Cabinet Office were unable to provide an estimate of their costs.

Conclusion

4.6 As in the 2011 data matching pilot, we are unable to provide a definitive overall cost for the pilot or comment on the likely costs should data mining be rolled out further. This is because we do not have final costs for all pilot areas or costs for any of the central government work. GDS spent a lot of time on this project and if that was included in the total spend the cost of the pilot is likely to increase significantly.

5 National data mining – judging success

5.1 This chapter summarises our findings in relation to the national databases used in the pilot. More details of the results submitted in relation to each database can be found in Appendices A-E.

Key points

- The evidence from this pilot suggests that data mining, as it was tested, is not a practical way of identifying unregistered electors.
- This is because, although the data returned to Electoral Registration Officers (EROs) did contain details of unregistered electors, it also contained significant numbers of existing electors, ineligible individuals and out of date information (where the individual was no longer resident at the given address).
- The reasons that so many existing electors and ineligible individuals were returned on the data include poor data specifications from Cabinet Office, currency restrictions not being tight enough and incomplete or poor quality addresses on some of the national databases.
- As a result of these issues, data mining is a resource intensive process. The volume of work involved in the pilot was far higher than expected for many pilot areas. The new registrations need to be considered in light of the time and resources required to achieve them.
- We are unable to draw clear conclusions on the cost benefit of data mining from this pilot, due to incomplete costs information.
- The pilot also does not provide for an assessment of the use of local data in comparison to national data, i.e. to what extent EROs could have found some of the individuals identified from the national data on their existing local data sources.
- In order for data mining to be of practical use to EROs, the data returned would need to contain many fewer names of registered or ineligible individuals and have significantly improved address information. Our recommendations consider how that could be achieved.
- Data mining, if rolled out nationally, would require a central organisation to take on the role of coordinating data transfer and processing the data (as Cabinet Office did for this pilot). Who that organisation would be is a key question for any data mining roll out.

Introduction

5.2 In judging the success or potential of national data mining there are two key questions:³⁹

- How effective is this data at identifying unregistered electors?
- Is data mining a cost effective way of registering new electors?

5.3 This chapter uses the results reported by the pilot areas and the issues raised by them in our interviews in order to reach a conclusion based on these two questions.

Identifying unregistered electors

5.4 The evidence from this pilot suggests that data mining, as it was tested, is not a practical way of identifying unregistered electors.

5.5 This is because, although the data returned to EROs did contain details of unregistered electors, it also contained significant numbers of existing electors, ineligible individuals and out of date information (where the individual was no longer resident at the given address).

5.6 We have not made, and could not make, a consistent assessment of what proportion of records returned to EROs represented unregistered people, resident at the given address. However, we have reported on what pilot areas found when they either checked the data against other data sources (e.g. their electoral registers) or when they received responses to their follow up work, as well as using their feedback on the processes and volume of work involved.

Addressing information

5.7 Pilot areas had serious concerns about the quality and consistency of address information on some of the databases used. The addresses held on each database are important as they form part of the matching process (an algorithm would ideally be able to recognise the same address on two different databases) and are also used to contact the potential new electors.

³⁹ Feedback from members of the public is also one of our evaluation criteria and, where relevant, we have included reference to that in the analysis presented here. However, while some pilot areas had expected a lot of objections to the principle of data sharing, in practice they found that they received relatively few. A couple of pilot areas even noted that members of the public were interested in and receptive to the aims of the scheme.

5.8 In relation to each of the national databases tested:

- **Department for Education (DfE):** addresses appeared to be more complete than those held in other national databases but a poor data specification from Cabinet Office meant that the format was inconsistent (see Appendix A for details).
- **Welsh Government, Department for Education and Skills (DfES):** the dataset used in this pilot (the Welsh annual pupil level school census) does not contain full addresses, only postcodes. For this reason, none of the pilot areas who received this data were able to do any work with it. There are therefore no further results to report for this database. Cabinet Office were aware of this limitation early in the process. It was always unlikely that pilot areas would have been able to make effective use of this data to identify potential new electors. The shortened pilot timetable made this even less likely.
- **Department for Work and Pensions (DWP):** there were numerous address issues including crucial address information being missing and extensive and confusing abbreviations.
- **Royal Mail:** the addresses appeared to be more complete and consistent than those held in other national databases. The extract used in the pilot did not contain UPRNs and two pilot areas (Renfrewshire VJB and Southwark) found it hard to match the records returned against their registers as a consequence.
- **Student Loans Company (SLC):** there seemed to be issues with the addresses on this data being incomplete. Only one pilot area (Greenwich) reported any usable results for this database and they found that nearly a third of the addresses were quite clearly incomplete, many of them unusable. SLC informed us that the addresses they provided to the Government Digital Service (GDS) were complete, so it seems that these issues may have arisen in the matching process, although we are unable to say for certain.

5.9 The weakness of some of the address information is likely to be due to the fact that on most of these databases, unlike the electoral registers, the address information is not central to the purpose for which the data is collected. The varying roles of address information are illustrated in Table 7 below.

5.10 Where the address on the registers and the address on another database are significantly different – due to misspelling, missing information, or different formatting – the matching process will not recognise the addresses. This record would then be returned to the ERO as a ‘mined’ record i.e. a potential new elector.

Table 7: Addressing on national databases

| Database | Role of address information | Address standardisation within the database |
|--------------------|--|--|
| Electoral register | Residence at address is required as part of eligibility to register. | Largely standardised within each register. Most EROs use UPRNs. |
| DfE | Gathered from schools, which hold addresses for correspondence purposes. | No standardisation; records received from multiple sources all using different address formats. |
| DfES | Gathered from schools, which hold addresses for correspondence purposes. DfES use home postcodes for some analysis but not for primary purpose of allocating school funding. | Postcode format is standardised (full addresses are not held). |
| DWP-CIS | Variable, depending on which source database the individual's details are drawn from. Some addresses used for correspondence, others not used. | Most source databases use standardised address formats although users are able to override this. DWP has been adding UPRNs for use in confirmation. Approx. 90% of records now have UPRNs. |
| SLC | One address used for correspondence; individual may enter another address which is not used. | Limited standardisation e.g. postcode must be within specified field. |
| Royal Mail | Previous address is used for identity checks; forwarding address is used for correspondence. | Completely standardised within the database. Royal Mail does not use UPRNs, although Transactis hold separate address files with UPRNs. ⁴⁰ |

5.11 It is likely that failed address matches accounted for a substantial proportion of the total number of records returned. GDS were unable to provide a breakdown of how many of the records returned to pilot areas had failed address matching. However, Greenwich provided some useful data. They looked at a sample of 170 SLC records which

⁴⁰ Transactis are Royal Mail's preferred data matching partner and conducted the data matching for this pilot.

had incomplete address information and were able to find 97% of these individuals on the register.

- 5.12 Addresses which are missing crucial information also mean that the ERO cannot use this address to write to the individual concerned, negating the point of providing this data.
- 5.13 Finally, on some of the databases used in the pilot, the address provided for an individual appears to be a correspondence address and the named individual does not reside at the property. Pilot areas reported this issue mostly in relation to individuals identified on the Royal Mail database, where an individual may choose to use a family or business address for their mail redirection if, for example, they are leaving the country or moving into temporary accommodation.

What was done in the pilot to try to solve this problem?

- 5.14 The issue of address standardisation between databases was also identified in our evaluation of the 2011 data matching pilot. One of our recommendations for any re-testing of national databases was that address compatibility problems needed to be mitigated, whether this was through the addition of UPRNs, the development of a national address file or through cleansing the data before matching.⁴¹
- 5.15 Since 2011, UPRNs have been added to around 90% of the DWP records, but not to any of the other databases.⁴²
- 5.16 Some pilot areas reported that address standardisation seemed to have been attempted but was not entirely successful. For example, all instances of 'Court' had been replaced with the abbreviated 'Ct', meaning that 'Harcourt Road' became 'Harct Rd'. This issue was mostly reported in relation to the DWP database, with some pilot areas reporting the majority of the records received from this database suffered from this issue. These abbreviations are part of the standardisation process in the DWP algorithm (as agreed with Cabinet Office), but had the unforeseen consequences of abbreviating incorrect parts of the addresses. It is likely that this issue could be resolved if there was any further testing of the DWP database for data mining.

What could be done in the future to solve this problem?

- 5.17 **The full addition of UPRNs to national databases used for data mining.** UPRNs are already used on many electoral registers and would

⁴¹ The Electoral Commission, *Data matching schemes to improve accuracy and completeness of the electoral registers – evaluation report* (March 2012) http://www.electoralcommission.org.uk/_data/assets/pdf_file/0010/146836/Data-matching-pilot-evaluation.pdf

⁴² This was part of an initiative undertaken by DWP for a range of purposes, not solely related to these pilots or to IER. Transactis hold UPRNs although Royal Mail does not, but the UPRNs were not used in the matching exercise for this pilot.

allow for the most straightforward matching process, with the algorithm having to match only a 12-digit code rather than several address lines. Coverage would need to be effectively universal to solve all of the problems. For example, given that the DWP now contains UPRNs, some pilot areas queried how much of the address information coming from this database could be poor quality. This is likely to be an inherent issue with searching for mismatched records – in the absence of universal UPRNs, the results will contain addresses which have failed to match.

- 5.18 Another advantage of using UPRNs, even where the underlying addressing information has not been cleansed, is that it could potentially enable the records to be returned with the addresses as stored on the electoral registers rather than on the databases. It could then be more practical for EROs to derive a list of people to write to or canvass.
- 5.19 Realistically, it seems unlikely that many other data holding organisations will chose to integrate UPRNs for their own purposes, because of the lower priority attached to addressing on these databases compared to the electoral registers.
- 5.20 **Improved standardisation of addresses between the registers and the databases.** Before the registers and databases are matched, there could be additional and more sophisticated cleansing and standardisation of data, i.e. ensuring that address formats are identical between the two datasets. However, this would mean a continued, significant central processing element in the data mining process.

Data currency

- 5.21 Pilot areas reported concerns about the currency of some of the data provided to them. Currency issues tended to broadly fall into two groups: records that were out of date due to home movement and records related to deceased people.
- 5.22 In relation to each of the national databases tested:
- **DfE:** the currency of this data appeared to be good.
 - **DWP:** a substantial number of the records returned to pilot areas related to individuals who were no longer resident.
 - **Royal Mail:** a substantial number of the records returned related to individuals who were no longer resident. This data also included deceased people, up to 1% of the records returned in one pilot area.
 - **SLC:** only one pilot area submitted usable results, but these indicate that a substantial number of the records returned related to individuals who were no longer resident. The timing of the pilot may have affected the currency of this data, as many students move at the end of the academic year.

5.23 Being provided with data which is not current creates two issues for EROs:

- the risk of writing to people who are no longer resident and therefore not eligible to register at that address
- the greater the volume of these records in the data provided, the more work the ERO and their staff are carrying out with no possibility of registering a new elector

5.24 In relation to the risk of contacting people who are not resident, this is partly a question of the perceived security of the system as the current residents may feel it is not appropriate to receive an invitation to register addressed to someone else.

5.25 There are also obviously concerns about sending letters to deceased people which will potentially be received by family members. Some areas reported receiving a lot of phone calls and complaints in response to inadvertently sending registration forms to deceased people, sometimes from family members who were very upset at receiving these. In a couple of cases, the ERO had already removed the deceased person from the register, having previously been informed of their death.

5.26 In order to mitigate these risks, pilot areas needed to undertake a lot of work cleaning and checking the data returned.

What was done in the pilot to try to solve this problem?

5.27 Currency restrictions were placed on all datasets used in the pilot. On the home movers option, for example, only records updated within the past 12 months on the DWP database and redirections set up within the past two years for Royal Mail were included. However, some pilot areas said that records going back this far were not current enough for the purpose of identifying highly mobile populations, and the data should be restricted to records updated more recently, for example within the past three or six months.

5.28 This is supported by evidence provided by four pilot areas who were able to separate their results by the 'age' of the DWP record. This shows that the numbers of new registrations are noticeably higher for records which have been updated in the past three months.

Table 8: DWP home movers data – new registrations by date of record update

| | Record last updated within... | | | |
|---|--------------------------------------|-------------------|-------------------|--------------------|
| | 0-3 months | 3-6 months | 6-9 months | 9-12 months |
| Total new registrations (% from sample) | 16% | 8% | 4% | 3% |

5.29 Pilot areas were not provided with the start date of the redirections on the Royal Mail data so we are unable to say for certain that this data would show the same pattern. It may be that, due to mail redirections being set up in advance of a move, the Royal Mail data stays current for longer (as there is likely to be a lag before DWP are informed of an address change).

5.30 For attainers, the DfE database extract was restricted to the most recent school census and this does appear to have been largely effective. There were a couple of comments from pilot areas that some of the DWP data used to try to identify attainers seemed to be years out of date. It appears that the data field used in the pilot to indicate whether the address has been updated in the past year may in fact sometimes indicate that the address has been re-submitted to the system, even though it has not changed.

What could be done in the future to solve this problem?

5.31 **Imposing a tighter restriction on record updates.** Evidence from some of the pilot areas looking at the home movers data suggests that a tighter restriction would improve the currency and could lead to a higher number of new registrations. DWP records with address changes in the previous three months resulted in higher numbers of new registrations in the four areas that compiled this data. We do not have comparable data for Royal Mail and any further testing of this data should provide EROs with the start date of the redirection in order to allow for this analysis.

5.32 The Royal Mail redirection service can be specifically used to re-direct the post of deceased individuals. Royal Mail has confirmed that deceased people can be excluded from the data provided in future, if requested.

5.33 If data mining was rolled out nationally, the issue of the data being out of date could be mitigated but not eliminated and EROs would have to reach their own views on the risks. Feedback from the pilot areas was mixed with some prepared to risk writing out to 'incorrect' names while others were more cautious and concerned. However, all of the concerned areas indicated that the time they spent cleaning and checking the data was unsustainable.

Existing electors

5.34 Ideally, the records returned through data mining would be only those who are not already on the register. In practice, many pilot areas found that a significant number of the returned records were in fact existing electors.

5.35 In relation to each of the national databases tested:

- **DfE:** we have limited results for this database but one pilot area found nearly 20% of the records related to existing electors.

- **DWP:** on average around 8-9% of the records related to existing electors, although in some pilot areas (particularly those who conducted more extensive checks against their registers) the figure is much higher.
- **Royal Mail:** on average around 13% of the records related to existing electors, although again it is much higher in some areas.
- **SLC:** we have limited results for this database but one pilot area found around 25% of the records related to existing electors.

5.36 The significant number of existing electors being returned is the result of two elements of the matching process:

- failure to match addresses (due to the issues discussed above)
- failure to match individuals' names (possibly due to abbreviation or mis-spelling)

5.37 It is a poor use of resources for EROs to send registration forms to existing electors and then have to process their responses (whether this is a completed form or a phone call to complain). Checking the names returned from the central databases against the register is a time-consuming process. The central matching process uses sophisticated and refined algorithms and so, in order to add value, it is likely that any further checks or matching at local level would need to be conducted manually. In addition, a couple of the pilot areas felt strongly that, if data mining was rolled out, it should be on the basis that EROs would not need to do any further checks on the data they received.

What was done in the pilot to try to solve this problem?

5.38 The matching algorithm used in this pilot was sophisticated and well developed, particularly in comparison to that used in the 2011 pilot. However, it was known that the returned data would still be likely to include existing electors.

What could be done in the future to solve this problem?

5.39 The suggested improvements set out above, in relation to address matching, would assist with this issue.

5.40 There are potentially improvements that could be made to the matching process in relation to names, but without clerical or 'eye-ball' checking, it is inevitable that some incorrect mis-matches will remain.

5.41 Given this, and the impracticalities associated with improving some aspects of address matching, it is likely that data mining would continue to return details of existing electors.

Ineligible individuals

5.42 The records returned in this pilot contained significant numbers of ineligible individuals, whether because of nationality or age.

5.43 In relation to each of the national databases tested:

- **DfE:** Cabinet Office specified an age range (16-18 year olds) rather than a date of birth range, which would have been more accurate. The data therefore contained individuals too young to be registered as attainers, as not all 16 year olds will be eligible.
- **DWP:** in some of the London boroughs in the pilot, this data contained significant numbers of foreign nationals.
- **Royal Mail:** this dataset contained a number of individuals below the age for registering to vote.⁴³ Many of the pilot areas were unhappy about the fact that they had been provided with such data. In addition, in some of the London boroughs in the pilot, this data contained significant numbers of foreign nationals.
- **SLC:** this dataset should contain relatively few foreign nationals. Only one pilot area submitted usable results and they found 5% of the sample were foreign nationals.

5.44 The issue of national databases returning names of foreign nationals is more of a concern for some areas than others. Unsurprisingly, areas with higher populations of non-EU nationals reported higher proportions of ineligible individuals being returned from the national databases. They were also more concerned about writing to foreign nationals to invite them to register. A couple of pilot areas said they would want to receive nationality information on the records returned.

5.45 Some of the returned data also included details of individuals who were below the age for inclusion on the register. Writing out to these individuals to invite them to register is embarrassing, as more than one administrator involved in the pilot commented. It is more likely to pose reputational risks than the risk of them actually becoming registered, but it also creates additional work for administrators in sending registration forms and processing responses.

What was done in the pilot to try to solve this problem?

5.46 It was known at the start of the pilot that there was no way to exclude ineligible nationalities from the mined results as none of the databases involved collect or record nationality information.

⁴³ The mail redirection form asks for the details of all members of the household who are moving. Not all of the members of a household will be eligible to vote.

5.47 For age, a date of birth cut off was applied to the DWP database in order to exclude those who were too young to be invited to register. This appears to have been largely successful at eliminating ineligible records on that database. No date of birth cut off was stipulated by Cabinet Office for the data requested from Royal Mail.

What could be done in the future to solve this problem?

5.48 The question of how to identify an elector's nationality when they apply to register is an issue beyond data mining. This has also been considered as part of the overall implementation of IER but no mechanism for checking nationality has been identified. This means that if data mining were to be rolled out, EROs would have to accept the risk that they are likely to receive ineligible foreign nationals in the results. As mentioned above, this is more of an issue for some areas than others, although in the pilot views were mixed across all types of area on how acceptable the risk was.

5.49 From discussions with Royal Mail, we understand that if Royal Mail data is included in any further data mining work, records relating to individuals below the age of registering to vote can be excluded from the information returned to EROs.

Cost effectiveness of data mining

5.50 The aim of this pilot was to see if data mining was an effective way of improving the completeness of the electoral registers, by identifying potential new electors who could subsequently be registered.

5.51 In order to answer this question, we would need to assess the cost benefit of data mining by, for example, calculating the cost per new elector registered. However, we are unable to do this as Cabinet Office could not provide details of their expenditure on the pilot. As they managed the process and conducted much of the matching and data processing, their costs could be significant and are crucial in reaching any realistic assessment of cost effectiveness.

5.52 In addition, we would need to assess the added value of providing EROs with access to national data compared to the data which they already have access to locally.

5.53 We do know the numbers of new registrations achieved by pilot areas vary widely, as set out below. These are quoted as percentages of the sample in order to be comparable between pilot areas, however in many cases, the actual numbers of new electors are low – on average around 300 individuals.

5.54 In relation to each of the national databases tested:

- **DfE:** new registrations ranged from 2% to 11% of the sample worked with, with an average of 9% (although we only have results from three pilot areas).
- **DWP:** new registrations ranged from 3% to 31% of the sample worked with, with an average of 8%.
- **Royal Mail:** new registrations ranged from 2% to 31% of the sample worked with, with an average of 8%.
- **SLC:** we only have usable results from one pilot area, where 4% of the sample worked with resulted in new registrations.

5.55 However, the level of new registrations does not provide a clear assessment of the potential of data mining. The registrations achieved were partly the result of the approach taken to data checking and follow up work. This varied across pilot areas but was generally limited to a single letter. The new registrations are therefore likely to be on the lower end of what could be achieved.

5.56 In addition, the delays in the pilot timetable mean that not all of the registrations achieved could be reported here (as responses will have continued to come in up to and following publication).

5.57 Part of the variation is also likely to be due to demographics. Some pilot areas had a higher response rate to their letters than others and in areas with higher levels of population mobility, it is likely that the data becomes out of date more quickly. From the data available, however, we cannot make any clear assessment of the impact of demographics on response rates.

5.58 Taken on their own, some of the numbers of new registrations appear reasonable (given the general lack of personal canvassing in the pilot, we have taken a 10% registration rate as reasonable). However, overall the numbers of new registrations are low in light of the time and resources spent to achieve them. Feedback from the pilot areas supports this assessment.

5.59 There is no evidence from this pilot that data mining is more or less effective for attainners or home movers. The average number of new registrations for both of these target groups is the same, at around 9% of the sample pilot areas worked with.

5.60 Data mining does not seem to have been as effective for students. This is largely because SLC was the only database which actually identified students and, as discussed above, there were issues with the addresses on the data received by EROs being incomplete. Cabinet Office also included data from DfE (in England) and DfES (in Wales) on the students option. However, neither of these datasets were suitable for identifying unregistered students: DfE because it does not cover further education colleges and DfES because it does not have full address

information, as discussed above. Cabinet Office's decision to use DfE data to try to identify students suggests a lack of understanding of this dataset.

- 5.61 Table 9 sets out the range of new registrations for each of the national databases as well as summarising our findings on the addressing and currency of the data and the extent to which it includes existing electors or ineligible individuals.

Time and resources

- 5.62 The issues with the databases discussed above added to the time involved centrally to match and clean data before it was returned to EROs.
- 5.63 Even if these issues were mitigated, there would still be a need for an organisation to take on the role of coordinating data transfer and processing data (the role that Cabinet Office took in this pilot). Who that organisation would be is a key question for any data mining roll out.
- 5.64 The need for a central coordinating body is key as some data holding organisations, such as DfE, do not conduct the matching process themselves. However, the requirement is wider than this and includes the management of relationships between national data holding organisations and local EROs. The alternative is requiring, for example, DWP to deal directly with individual data requests from 380 EROs.
- 5.65 This central organisation could be the eventual system owner for the IER Digital Service, which will manage the 'verification process' – the checking of electors' personal identifiers between EROs and DWP. However, in contrast to verification, data mining is unlikely to be as automated a process.
- 5.66 The issues with the databases also increased the workload falling on EROs when they received the data. Many of the pilot areas, including some recording relatively high numbers of new registrations, were clear in their feedback to us that the effort they had to put into the process was not sustainable and that they would not want to use this data, in this way, in the future.
- 5.67 In order for data mining to become part of 'business as usual' for EROs, the data returned would need to contain many fewer names of registered or ineligible individuals and have significantly improved address information. It would not therefore be prudent to invest resources in providing all EROs with access to this data, unless solutions to these issues can be identified, tested and implemented.

Table 9: National databases – summary of results

| Database | New registrations (% of sample) | Addressing | Currency | Existing electors | Ineligible individuals |
|-----------------|--|---|---|--|---|
| DfE | 2% - 11% | Limited results, but indications that addresses are of fairly good quality but inconsistent format. | Good. Relatively low numbers of individuals found to be no longer resident. | Limited data, but indications that relatively high proportions already registered. | 1%-8% ineligible; varies by demographics of area. |
| DWP | 3% - 29% | Incomplete and poor quality addresses. | Variable. Even when currency restrictions applied, seems to be around 9% no longer resident. | Around 9% of individuals on data are already registered. | 1%-5% ineligible; varies by demographics of area. |
| SLC | 0% - 4% | Incomplete and unusable addresses on data received by EROs (although not on the source data). | Variable. Limited results, but indications that relatively high proportions will no longer be resident. | Limited data, but indications that relatively high proportions already registered. | Limited results, but indications that low levels of ineligible individuals on data. |
| Royal Mail | 2% - 31% | Seems better quality than DWP but does not contain UPRNs. | Variable. Around 10% of individuals are no longer resident. Data identifies deceased individuals. | Relatively high proportions already registered – around 13% | Data included individuals below the age for registering to vote. Mail can also be redirected to correspondence addresses so individual is not necessarily resident there. |

- 5.68 Encouragingly, in contrast to the 2011 data matching pilot, we received little feedback about EROs and their staff not having the skills required to undertake this work. This appears to be the result of better processes employed in this pilot, such as only returning mis-matched names (in 2011 the areas received all the records including matches), as well as the fact that the majority of pilot areas had participated in a previous data matching pilot.
- 5.69 However, as in 2011, the workload involved was high because of the complexity, quality and volume of the data received, with time spent understanding the data, checking the records against the register, preparing and sending registration forms and processing responses (completed forms or phone calls).
- 5.70 In addition, some of the areas who have participated in a previous data matching pilot still struggled to some extent to understand the data. For EROs with no previous involvement in data mining, there could be a steep learning curve and any roll out would need to be managed in a way that is sympathetic to differing local needs.
- 5.71 Although many pilot areas reported spending more time than anticipated on the pilot, often for few new registrations, we are unable to produce a cost-benefit analysis of data mining. This is because we do not have any information on the costs incurred by Cabinet Office, whose input was a crucial part of the process (see Chapter 4 for details).

Conclusion

- 5.72 National databases can provide EROs with the details of significant numbers of currently unregistered, potential electors. However, the practical difficulties in using this data are wide-ranging and affect all of the national databases used in the pilot, albeit to varying degrees.
- 5.73 Addressing issues are the most significant, due to the importance of accurate and consistent address information in the matching process and for contacting the individuals identified. The variable quality and formatting of addresses reflects the differing roles of addresses on the different databases. In many cases the address information is less important than it is on the electoral registers.
- 5.74 Nearly all of the pilot areas reported spending significantly more time on the pilot than they had expected. This was primarily due to the volume and quality of data received. This input of time and resources clearly needs to be taken into account when assessing the numbers of new registrations achieved. While in some areas reasonable results were achieved, this was generally the result of an unsustainable level of input.
- 5.75 Importantly, the pilot did not provide for an assessment of local data in comparison to national data, i.e. to what extent areas would have found some of the names provided in the pilot on their existing local data

sources. In most cases, it would be more straightforward for an ERO to access a local data source as they tend to work to similar standards of addressing within the local authority.

- 5.76 We are also unable to draw clear conclusions on the cost benefit of data mining since we do not have any costs for the time Cabinet Office spent on the pilot.
- 5.77 The DfES database is not suitable for the purposes of data mining, as it does not contain full addresses.
- 5.78 Overall, although there remains potential in the concept of data mining, until many of the issues outlined above are rectified, national data mining is unlikely to be of use to EROs. The evidence from this pilot therefore does not justify the national roll out of data mining. Our recommendations set out what we consider to be the most important changes required for any future testing of data mining.

6 County data mining – judging success

6.1 This chapter sets out our conclusions on the county data mining. Further details of the results from the county data mining can be found in Appendix F.

Key points

- This pilot did not try to assess the usefulness of access to county council data. Rather it was set up to provide qualitative feedback on the barriers and issues that would be faced if Electoral Registration Officers (EROs) in lower tier authorities tried to access the data held by an upper tier.
- There is therefore nothing in the findings to challenge the assumption that it would be sensible to equalise unitary authority and lower-tier authority EROs' access to data.
- However, it took a long time and a great deal of effort to establish the data sharing arrangements between the pilot areas and the county councils. It is clear that, even if legal access was granted to lower tier EROs, they would still need to invest time in securing access to, and learning how to best utilise, the county data.
- Importantly, EROs would need to assess the cost effectiveness of accessing the data that became available to them – exactly as they should do with the local data they currently have access to.

6.2 The aim of this part of the pilot was to explore the potential issues and barriers that would be faced when sharing data between two tiers of local government. As such there were no real success criteria for this part of the pilot. We have considered two main questions:

- What lessons can be learned from setting up these arrangements?
- Could this data help to identify unregistered electors?

What lessons can be learned from setting up these arrangements?

6.3 A significant amount of work was required to put the data sharing arrangements in place and the lower tier authorities had the support and influence of Cabinet Office in this pilot.

- 6.4 Only one of the county councils responded to our request for an interview, so we do not know the specific issues that caused delays in each case. However, it seems that at least some of the county councils had significant concerns about sharing data with their lower tier authority, and in some cases it was difficult to identify an individual or team within the county who would take responsibility for finalising the agreements.
- 6.5 This suggests that, in the event of a change in the law opening up access to upper tier data, many lower tiers would still need to put time into securing the data. And, even if the law was changed, lower tier authorities would still need upper tier cooperation to ensure that the data was provided in a usable format.
- 6.6 Based on the experiences in this pilot, the issues that lower tier authorities may encounter include:
- A lack of pre-existing relationships between the upper and lower tier authorities and no experience of sharing data between themselves in this way.
 - Few direct benefits or incentives for upper tier authorities to participate since they do not have any responsibility for electoral registration.
 - Concerns over data protection and how the lower tier will store and use the upper tier's data.
- 6.7 However, more positively the pilot has shown that data can be shared between an upper and lower tier.
- 6.8 One of the findings from this work is that personal relationships are key. This is perhaps unsurprising in the sense that trust is an important element in securing access to data. The obvious downside to this is that staff turnover in either authority could have an impact on arrangements, although that could be expected to be less of an issue over time, as the data sharing arrangements 'bed-in'.

Could this data help to identify unregistered individuals?

- 6.9 At its simplest, this data can be accessed by EROs in unitary authorities and it is peculiar that EROs in lower tier authorities do not have the same right of access.
- 6.10 However, from the outcomes of this pilot we are unable to say that access to the type of data tested (generally education data) would be useful to lower-tier EROs – only three out of the four pilot areas looking at the county data submitted results.

- 6.11 Each pilot area received a different dataset. As would be expected, the quality and suitability of the data they received varied.
- 6.12 One pilot area told us that the data seemed to be current and accurate, while another felt that it was very out-of-date, receiving responses to their letters informing them that the individual had moved out of the property many years previously.
- 6.13 Of the three pilot areas who provided results, two reported finding no ineligible individuals on the county data, with the third finding only a very small number. This may reflect the demographics of these areas, with a smaller non-EU population than other parts of Great Britain.
- 6.14 Across the three areas, the number of new registrations achieved was 12% of the samples worked with.

Conclusions

- 6.15 It is difficult to draw firm conclusions from a small number of pilot areas but there is nothing in the findings to suggest that it would not be sensible to equalise unitary authority and lower tier authority EROs' access to this type of data.
- 6.16 It is clear that if access was granted EROs would still need to invest time in securing access to, and learning how to best utilise, the county data.
- 6.17 Importantly, EROs would still need to assess the cost benefits of accessing the data that became available to them – exactly as they should do with the local data they currently have access to. As part of that they should satisfy themselves of the robustness of any data source they were granted access to and its suitability for helping to maintain the register.

7 Recommendations

7.1 This chapter sets out our recommendations based on our evaluation.

National data mining

The findings from this pilot do not justify the national roll out of data mining. The concept of using national data to assist EROs may still have potential, but **data mining should not be implemented without further testing of the databases and processes.**

Data mining would require a central organisation to be responsible for managing the connection between national data holding organisations and undertaking data processing work. Cabinet Office undertook this role for this pilot.

The need for a central coordinating body is key as some data holding organisations, such as DfE, do not conduct the matching process themselves. However, the requirement is wider than this and includes the management of relationships between national data holding organisations and local EROs. The alternative is requiring, for example, DWP to deal directly with individual data requests from 380 EROs.

This central organisation could be the eventual system owner for the IER Digital Service, which will manage the 'verification process' – the checking of electors' personal identifiers between EROs and DWP. However, in contrast to verification, data mining is unlikely to be as automated a process.

There should only be further data mining testing on the understanding and acceptance of the need for an ongoing central presence (and any related costs) in order to receive the data from EROs and the national organisation, match it and return the results to each local area.

Any further testing should also be considered in relation to the priority of the overall transition to IER. Plans should therefore take into account the capacity of all the organisations and individuals required to test data mining, specifically to ensure that any testing would not adversely affect their existing commitments to delivering IER.

In addition, there were numerous issues in this pilot with the communication and support provided by Cabinet Office. It is important that **Cabinet Office considers what lessons can be drawn from this pilot, particularly in terms of engagement with EROs, for the wider implementation of IER.**

If further testing is undertaken then, in relation to specific databases:

- **There would be merit in re-testing the Department of Education database** as there were fewer issues than for the other databases and limited results were returned. However:
 - it would be sensible to explore the approach to addressing on this database ahead of any full pilot
 - 16 year olds who are under attainer age should be excluded from the data returned to EROs
- **The Welsh Department for Education and Skills database is not suitable for the purposes of data mining and should not be tested again.**
- **The Department of Work and Pensions database should only be included in further testing if:**
 - **full integration of UPRNs is completed** (i.e. all records have UPRNs)
 - **the record currency can be restricted to those with address changes within the past three months, at most**
- **The Royal Mail database should only be included in further testing if:**
 - **the names of individuals below the age for registering to vote are excluded from the data, which Royal Mail has confirmed it will do**
 - **the record currency can be restricted further and the data shared includes the start date of the redirection**
- **The Student Loans Company database should only be included in further testing if:**
 - **the addressing issues experienced in this pilot can be resolved**
 - **testing takes place during October - November or January - February rather than at the end of the academic year**

In relation to any further testing in general:

- **There needs to be a clear understanding of the databases being accessed and a clear data specification provided to the data holding organisations** (based on the requirements of the pilot). This would be particularly important for any new databases included.
- **For any new database proposed for data mining testing, their approach to addressing should be assessed in advance of the pilot.** There would be limited value at this point in testing a database which lacks UPRNs and has poor addressing information.

- **For any database tested, the potential for returning records to EROs with the original register address attached should be explored** (rather than the address held on the national database). This could, where available, be achieved using UPRNs.
- **Any combination of databases needs to be less complex.** For example, in this pilot data from two databases was combined into one file, and pilot areas could receive more than one file. Many areas could not clearly report on the results and this made evaluating the pilot more difficult. **For future testing, only one file with data from one database should be provided to each ERO.**
- **There should be mandatory checking of the national data provided against data held locally.** This would allow for an assessment of the added value to EROs of access to national data, as compared to local data which they already have access to. For example, if DfE data is included in the re-testing, it should be compared with locally held education data to assess whether the unregistered individuals identified on DfE could be identified using local data instead.
- **Cabinet Office need to ensure that they maintain good communication between themselves, the data holding organisations and EROs throughout the process,** including after data from the national databases has been returned to EROs.

County data mining

The results from this pilot do not show how useful it would be for EROs for lower tier authorities to have access to data held by an upper tier. However, **EROs in these authorities should be given the legal right of access to data held by upper tier authorities, to put them in position analogous to EROs in unitary authorities.**

EROs are responsible for deciding which local data they are prepared to use in maintaining their register. These decisions should be based on an assessment of the quality of the specific database to be used.

Appendix A – Department of Education

This appendix sets out our findings in relation to the Department for Education (DfE) database.

Data held by the Department for Education (DfE) was used to try to identify attainers and students (for England only). Seven pilot areas accessed DfE data: three for attainers only, one for students only and three for both. The Government Digital Service (GDS) combined DfE data with data from the DWP database on the attainers option and with data from the SLC database for students before sending the information to pilot areas.

However, the DfE data is not relevant for identifying further or higher education students, as the dataset does not cover further education colleges (or universities, although this was always known). Its inclusion was the result of a lack of understanding of the database on the part of Cabinet Office.⁴⁴

The data provided to pilot areas in both files covered school pupils aged 16-18. This created some complexities for the areas who looked at both the attainers and students options, as they received a large number of records providing information on the same people.⁴⁵

In considering the findings presented below it is important to note that the range of evidence is limited as not all areas carried out the type of checks that are discussed and not all areas reported clearly on new registrations achieved solely as a result of DfE data.

Quality of the data

Addressing

There seem to have been fewer issues with addresses being inaccurate or incomplete (e.g. missing addressing information) on the DfE database than

⁴⁴ DfE does hold data on some students at further education colleges but this does not include a full home address and so was not used in the pilot.

⁴⁵ Of the three areas which received DfE data on the attainers and students options, two removed some (but not all) of the duplicates before sending records and none were able to separate their results by database.

the other national databases.⁴⁶ However, addresses did tend to have quite inconsistent formats. This is to be expected as the database draws from a large number of sources, e.g. schools, and DfE do not apply any address standardisation as part of business as usual.⁴⁷

We understand that DfE would have been prepared to attempt to standardise addresses in the extract used in this pilot, e.g. to ensure all street names appeared in a single field, however the data specification provided was unsuitable.⁴⁸

Currency

Greenwich found that around 4% of the individuals identified on the DfE file no longer lived in the property. We do not have figures for Sunderland. Coventry and Wolverhampton reported finding around 2% of individuals had already moved out (although this was for individuals identified on more than one databases).

The DfE extract used for the pilot only contained records from the most recent school census. The number of individuals found to have already moved from the address given on the dataset does seem to be lower than for other databases, even for the same group of electors. For example, some of the attainers identified on the DWP database were found to have moved away 10 or more years ago (see Appendix C). This suggests that the currency restriction on the DfE database was fairly successful, although we have insufficient data from which to draw a firm conclusion.

Existing electors

By checking the records received against their register, Greenwich found that just under 20% were existing electors. Sunderland, however, found only 2% of the records related to existing electors. This difference may be due to the method or extent of checks against the register.

Coventry and Wolverhampton reported that around 8-9% of the individuals identified were existing electors, with Tower Hamlets finding over a third to be

⁴⁶ Greenwich, for example, found only five incomplete addresses out of 3,524 DfE records, compared to 776 incomplete addresses out of 2,754 SLC records. These figures do not represent a comprehensive logging/analysis of all address issues in file, but rather records with address issues that were spotted as Greenwich worked with the data. As they did roughly the same amount of work with DfE and SLC records these figures indicate that there was a far higher proportion of address issues in the SLC file.

⁴⁷ The DfE database allows for up to seven address lines to be entered; the number that is actually used will vary depending on the source of the data.

⁴⁸ The specification asked them to reduce their seven address lines into three. In the 2011 pilot DfE were asked, and were able to, reduce into five lines, but three would have been impractical and too much work. As a result no standardisation took place.

already on the register. However, these sets of figures cover individuals returned from the DWP and SLC databases as well as DfE records.

The relatively high percentage found on the register in some areas is, in part at least, likely to be caused by difficulties in address matching. A clearer and more realistic data specification for the DfE could potentially have helped to reduce the number of individuals returned who were already registered.

Ineligible individuals

Greenwich found that around 8% of their sample were foreign nationals who were ineligible to register.⁴⁹ The proportion found to be ineligible in Coventry, Tower Hamlets and Wolverhampton was less than 1%.

The variation in these findings, while surprising (given that Tower Hamlets has a similar proportion of non-EU nationals to Greenwich) is likely to be due to the extent of checks made against the register, the methods used in conducting these checks, the approach adopted by the ERO (i.e. using this as an opportunity to seek further information about nationality) and/or the availability of data held locally to assess nationality.

Due to an inappropriate data specification, the data included some individuals who were too young to be registered as attainers. This is because Cabinet Office specified an age range (16 – 18 year olds) rather than a date of birth range, which would have been more accurate as not all 16 year olds will be eligible as attainers.⁵⁰ This issue caused some pilot areas additional unexpected work in sorting and excluding these individuals from their sample.

Other issues

The use of an age range rather than a date of birth range meant that the attainers data also included individuals who were already 18, so entitled to be registered as full electors rather than attainers. While this data is still useful to EROs, it caused some complexities in reporting results for this pilot as different pilot areas adopted different approaches to this group of individuals, with some including them in their pilot follow up work, some contacting them outside of the pilot processes and some not conducting any follow up work.

⁴⁹ Much of this was picked up during the checks against the register, where administrators could see that the family name matched the one listed for that property, with a previous canvass response indicating that they were all foreign nationals.

⁵⁰ An attainer is an individual who will turn 18 before the end of a 12 month period starting from the next 1 December after the application is made i.e. if an application is made in spring 2013, the applicant will be eligible as an attainer if they turn 18 any time before December 2014.

New registrations

We only have results from three areas in relation to the DfE data, since the other areas were unable to separate their results by database.⁵¹

Due to poor communication from Cabinet Office and inaccurate data labels, in Harrow looked only at a subset of the DfE data – the records that were found only on the DfE data and not on both DfE and Department for Work and Pensions data. It is likely that by excluding records found on both databases, Harrow followed up the less accurate or current DfE records, which perhaps explains their low registration rate.

Coventry, Tower Hamlets and Wolverhampton were unable to separate the results for the different databases they accessed in their reports. All three areas received both the attainers and students files which means that, as discussed above, they will have received a large number of duplicate records.⁵² The numbers of new registrations in these areas range from 5% to 15% of the total sample, but it is not possible to know which of these registrations came from records identified on DfE data, and which from records identified on DWP or SLC data (or on a combination of databases).

Rushmoor did not separate out databases in their reporting. Although figures provided by GDS indicate that most of the records returned from DfE were also found on the DWP database. Rushmoor's results are therefore reported in Appendix C on the DWP database as it is likely that these figures also contain the results for 14 records found solely on DfE.

Conclusion

The currency of data on the DfE database appears to have been good. However, while the addresses on the database seem to have also been of reasonably good quality, inconsistencies in their formatting seemed to create problems in the matching process, as shown by the high proportions of existing electors returned in the data.

We have limited quantitative data to draw on in relation to the DfE dataset. The number of new registrations in the three areas that reported results specifically for DfE data is broadly in line with the average on other datasets – 9% in total. However, these new registrations need to be balanced against the amount of work involved (see Chapter 5).

⁵¹ Some pilot areas reported one set of figures for each option, combining results for individuals found on various databases e.g. reported one set of figures for the attainers file which covered individuals found on the DfE database, DWP database and on both databases. This was largely due to difficulties in interpreting the spreadsheet or in tracking the responses.

⁵² As the DfE data provided to pilot areas in both files covered the same individuals.

Table A1: New registrations from DfE data

| Area | Sample size ⁵³ | Number of registration forms sent | Number of new registrations | % new registrations (from total sample) |
|---------------|---|-----------------------------------|-----------------------------|---|
| Coventry | Results not separated by database ⁵⁴ | | | |
| Greenwich | 430 | 314 | 41 | 10% |
| Harrow | 584 | 519 | 13 | 2% |
| Rushmoor | Results reported in Appendix C under DWP data | | | |
| Sunderland | 2,266 | 2,219 | 257 | 11% |
| Tower Hamlets | Results not separated by database | | | |
| Wolverhampton | Results not separated by database | | | |
| Total | 3,280 | 3,052 | 311 | 9% |

Some of the issues identified in relation to the DfE data could have been avoided by better data specifications for address formatting and date of birth ranges. A better understanding of the dataset may also have prevented it being used to try to identify students, as it is not an appropriate dataset for this purpose.

There may be merit in further testing of the DfE database for the purpose of identifying attainers although this should only be done with an improved data specification. It would also be sensible to review the addressing used on the database further ahead of any full pilot.

⁵³ Differences in sampling mean results are not necessarily comparable between pilot areas. We have tried to make samples as consistent as possible by applying the following principles:

- If an area worked with the total number of records returned or selected a random sample, this is their sample size.
- Where areas wished to restrict follow-up work to the wards or postcodes they had originally planned to work with, the sample excludes records with addresses outside of these locations.
- On the attainers and county options, the sample excludes any records the pilot area deemed too old or too young to be included as attainers.
- If areas did not follow up certain records because of address or eligibility issues etc., these records have been added back into the sample where possible.

⁵⁴ Two databases were included for each option. Some pilot areas reported one set of figures for each option, combining results for individuals found on various databases.

Appendix B - Welsh Government, Department for Education and Skills

This appendix sets out our findings in relation to the Welsh Government's Department for Education and Skills Pupil Level Annual School Census database (DfES).

Data from DfES was provided to Ceredigion and Powys for the attainers option. Pembrokeshire was also supposed to receive this data but there were problems around the file transfer.⁵⁵

Powys received over 17,000 records from the DfES database, equivalent to approximately 15% of their total population. While GDS believe this error may have been caused by problems with the input file, the Welsh Government told us that they used the same data specification for each of the pilot areas.

The DfES dataset does not include full addresses, only postcodes. For this reason, neither Ceredigion nor Powys were able to undertake any follow up work on this data. There are therefore no further results to report.

Cabinet Office knew that the DfES data did not include full addresses from the early stages of the pilot. They thought that pilot areas might be able to combine the DfES data with other locally held data, or with the match results returned from DWP, in order to identify the full addresses. However, this did not prove feasible, partly because of the shortage of time and partly because it was always uncertain whether this approach would yield any usable information.

Conclusion

The DfES database is not suitable for the purposes of data mining because it does not contain full addresses, which are a minimum requirement in order to check records against the register and then contact potential new electors.

⁵⁵ Pembrokeshire believe that they did not receive this data, whereas Cabinet Office informed us that it had been sent. It appears that while Pembrokeshire did receive the disc containing the data they did not use it as an updated version of one of the files was quickly sent out by Cabinet Office by email. Pembrokeshire worked with data provided by email but as the update related to home movers data, believed that the attainers file had not been sent. The problem was not resolved before the data transfer window closed although we do not believe any data was lost between Cabinet Office and Pembrokeshire.

Appendix C - Department for Work and Pensions

This appendix sets out our findings in relation to the Department for Work and Pensions Customer Information System (DWP-CIS) database.

Data extracted from the DWP database was used to try to identify attainers and home movers. 18 pilot areas accessed DWP data: four for attainers only, nine for home movers only and five for both. The Government Digital Service (GDS) combined DWP data with DfE data for the attainers option and with Royal Mail data for the home movers option before sending the information to pilot areas.

Quality of the data

It is important when considering the findings below to remember that any concerns raised over the quality of addresses or currency of data do not undermine the principle of using the DWP data for confirmation. The results considered below are for non-matches between the registers and the DWP database and as such they are inherently more likely to encounter issues with records held by DWP.

Our assessment of confirmation remains the same – it is an accurate, effective process for retaining electors on the register during the transition to IER.

Addressing

Many of the pilot areas who used DWP data had extensive complaints about the format and quality of the addresses on the DWP records. Issues included:

- Crucial address information missing, such as the flat number, street number or even street name. Some addresses only referred to the general district or borough.
- Extensive abbreviations, for example the use of 'St' instead of 'Street'.⁵⁶ A few pilot areas commented that it looked like this had been done using a 'find and replace' tool, because the abbreviations had been applied to

⁵⁶ Wrexham estimated that 60% of their addresses had been abbreviated. Richmond also raised this issue but with regards to the data they received from Royal Mail. However, they were the only area to do so. It may be that they were the only area to notice that this was also happening with the Royal Mail data as they received substantially more Royal Mail records than other areas.

the name of the road as well. For example, pilot areas came across addresses such as 'Pk La' for Park Lane and 'Dunbla Rd' for Dunblane Road. The abbreviations are part of the standardisation process in the DWP algorithm (agreed with Cabinet Office), but had the unforeseen consequences of abbreviating incorrect parts of the addresses. It is likely that this issue could be resolved if there was any further testing of the DWP database for data mining.

- References to counties and districts which no longer exist, including references to pre-1996 counties in Wales. Some Welsh pilot areas reported that half or more of the DWP records contained these county names.
- Greenwich and Richmond reported that few of the records had UPRNs, and even those that were present were not all correct.⁵⁷

Some pilot areas spent a long time trying to unravel the data, work out what the correct addresses were and check them against their electoral register. This includes Greenwich, who despite doing a lot of time-consuming investigative work, still found that 9% of the DWP addresses were unusable. In addition, of the 1,623 letters they eventually sent out, 97 were returned as undeliverable (6% of the letters, 3% of their total sample).

Conversely, Renfrewshire found the DWP data fairly easy to work with. A very high proportion of their DWP records had UPRNs which enabled them to easily match the records against their register, using an automated process. Richmond also said that the DWP data was better quality than the Royal Mail data provided alongside, which is why they chose to do more follow up work for the records returned from DWP.

Currency

All DWP records used in the pilot should have had an address update in the last 12 months. However, some pilot areas noted being informed that some of the individuals they were writing to had moved out of that address many years before, on both the attainers and home movers options. On average, around 2% of individuals identified on the DWP home movers data were found to be no longer resident, although in Powys and Richmond, two of the areas who conducted extensive follow up work, this was 22% and 15% respectively.

It appears that the data field used in the pilot to indicate whether the address has been updated in the past year may in fact sometimes indicate that the address has been re-submitted to the system, even though it has not changed.

⁵⁷ There are some known issues with the assignment of UPRNs to the DWP-CIS data which DWP is addressing in advance of the confirmation dry run in July 2013 but which were still present for the data mining pilot.

Some pilot areas also queried how the data could include references to counties which had not existed in over 12 years. This appears to be because either the address was not actually updated when the database indicated it had been or because the county name may appear in address line 5 or 6 and while updates are provided to the initial address lines, they may never explicitly remove the reference to a county in a later field.

Despite these issues with the address updates, results from pilot areas who conducted this analysis suggest that in general the numbers of new registrations are higher for records which have been updated more recently.

Table C1: Registration rates by date of last record update

| Pilot area | Record last updated within... | | | |
|--------------|-------------------------------|------------|------------|-------------|
| | 0-3 months | 3-6 months | 6-9 months | 9-12 months |
| Greenwich | 10% | 3% | 2% | 2% |
| Lothian VJB | 9% | 5% | 7% | 2% |
| Powys | 42% | 28% | 13% | 11% |
| Wrexham | 13% | 8% | 4% | 3% |
| Total | 16% | 8% | 4% | 3% |

Evidence from this pilot indicates that a 12 month restriction was too broad. Any further testing of the DWP database should be restricted to records which have been updated within the past three months only.

A couple of areas also reported finding deceased individuals on their home movers data, either through checking against the previous versions of their register or in response to the follow up activity. However, pilot areas were mostly unable to provide any figures so we are unable to state definitively which database the individuals in question were identified on.

Existing electors

On average, pilot areas reported that around 8% of the individuals identified on the DWP data were existing electors.⁵⁸ In a couple of areas – Conwy and Greenwich – this figure is over 20% and in Greenwich, this is six times the rate of new registrations. Other than Powys, Richmond and Sunderland, all areas looking at DWP data found the proportion of existing electors was at

⁵⁸ This analysis excludes Rushmoor – they found 24% of their sample on the register but the extract sent to Cabinet Office for matching did not include some of the 16 year olds they had details for (i.e. individuals who were not quite attainers at the date of the canvass but whose information had been included on the canvass form).

least equal to, and often far higher than, the proportion of successful new registrations.

The fact that the data returned to pilot areas would include some existing electors was known from the results of the confirmation pilot. Ideally, for the purposes of data mining, this category would be as small as possible. It is likely that at least some of the existing electors who were returned as potential new electors were because of problems with address matching and also differences in how names are recorded between the database and the electoral registers.

Ineligible individuals

The proportion of individuals found to be ineligible (largely because of nationality issues) is 1% of the total sample or less, except for Greenwich and Richmond where the rate is around 4-5%.⁵⁹

It was known at the start of the pilot that there was no way to exclude ineligible nationalities from the mined results as none of the databases involved collect or record nationality information.

New registrations

The numbers of new registrations for individuals identified on the DWP data ranged from 5% to 29% for attainers and 3% to 26% for home movers, as shown in Tables C2 and C3 below.

Attainers

Due to poor communication from Cabinet Office and incorrect data labels, Harrow followed up only a subset of the DfE records and so did not do any work with the DWP data.

Coventry and Wolverhampton were unable to separate databases in their results for attainers. Coventry reported an 11% registration rate and Wolverhampton 9%, but it is not possible to know which of these registrations came from records identified on the DWP, DfE or both.

Similarly, Tower Hamlets did not separate their results by database and combined their results for both attainers and students. These results are discussed in relation to the DfE data (as this is included on both files) and are included on their profile in Appendix A.

⁵⁹ Some areas with high non-EU populations did not find a high proportion of ineligible individuals on the data. This is likely to be due to several reasons, including the extent of checks made against the register, the methods used in conducting these checks, the approach adopted by the ERO (i.e. using this as an opportunity to seek further information about nationality) or the availability of data held locally to assess nationality.

Table C2: New registrations from DWP data (attainers)

| Area | Sample size ⁶⁰ | Number of letters | Number of new registrations | % new registrations (from sample) |
|---------------|---|-------------------|-----------------------------|-----------------------------------|
| Ceredigion | 1,763 | 611 | 87 | 5% |
| Coventry | Results not separated by database ⁶¹ | | | |
| Harrow | No follow up conducted | | | |
| Pembrokeshire | Did not receive data ⁶² | | | |
| Powys | 391 | 377 | 120 | 31% |
| Rushmoor | 1,771 | 1,301 | 166 | 9% |
| Sunderland | 6,110 | 5,798 | 466 | 8% |
| Tower Hamlets | Results not separated by database | | | |
| Wolverhampton | Results not separated by database | | | |
| Total | 10,035 | 8,087 | 839 | 8% |

Rushmoor have not separated by database in their report. However their sample includes only 14 records found solely on DfE data. Their results are therefore reported here as they are more representative of the DWP data. Their sample is also inflated because the extract of their register used to match against the DWP data did not include all of the registered attainers. It is likely that the proportion of new registrations reported should be higher, although we cannot say by how much.

Home movers

The relatively high registration rate in Richmond is likely to be because they conducted more follow up work on the DWP data, including sending a reminder letter and canvassing non-responders.

We cannot explain why Powys was able to achieve such a high registration rate in relation to other areas. They did conduct some canvassing but within one postcode area only. Powys has the highest registration rate for all the databases and target groups they looked at. This could be due to the demographics in this area.

However, we have been unable to clarify some elements of their approach to sampling. We do know that Powys focused on particular target areas and removed addresses that fell outside these areas. It may be that they also

⁶⁰ See footnote 53 against Table A1 for details of how the samples have been calculated.

⁶¹ Two databases were included for each option. Some pilot areas reported one set of figures for each option, combining results for individuals found on various databases.

⁶² See footnote 55 above

Table C3: New registrations from DWP data (home movers)

| Area | Sample size | Number of letters | Number of new registrations | % new registrations |
|---------------------------------|-----------------------------------|-------------------|-----------------------------|---------------------|
| Ceredigion | 335 | 287 | 30 | 9% |
| Conwy | 3,144 | 2,438 | 357 | 11% |
| East Renfrewshire ⁶³ | 610 | 535 | 59 | 10% |
| Greenwich | 3,148 | 1,623 | 132 | 4% |
| Lothian VJB | 453 | 373 | 30 | 7% |
| Pembrokeshire | Results not separated by database | | | |
| Powys | 870 | 814 | 250 | 29% |
| Richmond-upon-Thames | 609 | 475 | 102 | 17% |
| South Ribble | No follow up work conducted | | | |
| Southwark | 4,729 | 4,729 | 161 | 3% |
| Sunderland | 13,187 | 12,720 | 1,067 | 8% |
| Tower Hamlets | 9,317 | 8,183 | 632 | 7% |
| Wigan | Results not separated by database | | | |
| Wrexham | 2,211 | 2,051 | 190 | 9% |
| Total⁶⁴ | 38,613 | 34,228 | 3,010 | 8% |

removed poorer quality addresses as part of this process which would have had a positive effect on the response rate.

Wigan was unable to separate the two home movers databases in their report, but they achieved a 13% registration rate overall.

Pembrokeshire reported a 34% registration rate from letters. They excluded any records where the address included an out-of-date county and possibly those with other address issues. We have not been able to fully clarify their approach with them and it is possible that they entirely excluded DWP records from their follow up work (i.e. only working with Royal Mail data).

⁶³ Renfrewshire VJB conducted checks against the register for all DWP records with UPRNs (2,469 records). They found 305 on the register (12%). They then took a random sample of around 20% of the records not on the register to write to. We have added back in the proportion of records they found on the register to their sample in order to provide comparable figures.

⁶⁴ Calculations exclude Pembrokeshire as they did not report a sample size

Tower Hamlets worked with the 'newest' DWP data, following up all of the records which had been updated in the past three months and some of the records which had been updated in the past six months.

Conclusion

DWP data was accessed by more pilot areas than any of the other national databases in the pilot, and that may be why there have been more issues identified in relation to it.

The problems with addresses were the most common issue raised. In their interviews with us, many pilot areas were negative about the usefulness of the DWP data because of these problems. The issues with extensive abbreviations were caused by the processing and matching of the data, but issues such as missing flat or house numbers seem to be more integral to the dataset. UPRN coverage is also not universal and addresses with missing UPRNs were always likely to be more prevalent in the data returned to pilots than in the overall database.

Currency of the data also appears to have been an issue, despite the restriction to records where there had been an address update in the last 12 months. Feedback from pilot areas suggested that a tighter restriction would be useful and evidence provided by some areas does suggest that focusing on records updated within the previous three months would yield better results.

The number of new registrations varied significantly between the pilot areas and it is not always clear what has created that variation. The rates of new registrations between the attainers and home movers options seem similar, although we have limited data for attainers. These new registrations need to be considered against the amount of work involved in using this data.

Any further data mining testing should only consider including DWP data if the issues set out above in relation to addressing and currency are resolved.

Appendix D - Royal Mail

This appendix sets out our findings in relation to the Royal Mail database.

Data held by Royal Mail was used to try to identify home movers. The Government Digital Services (GDS) combined it with data from the DWP database before sending the information to pilot areas.

Quality of the data

Addressing

Views on addressing generally seemed more positive than on the DWP home movers data, with a number of areas reporting that the addresses were better quality and more standardised. However, one area (Richmond) reported the Royal Mail data had the same issue with the extensive and confusing abbreviations. A couple (East Renfrewshire and Southwark) found the Royal Mail data difficult to work with because it did not include UPRNs, which would have allowed them to check the records more easily against their registers.⁶⁵

Currency

Only redirections set up within the past two years were included in the dataset for this pilot. Pilot areas had mixed views on the currency of the data. On average 10% of individuals were found to be no longer resident. Some pilot areas report being informed that the person identified had not lived in the property for many years. A few pilot areas were unhappy that the extract included records up to two years old, feeling that this was too long for this group of electors.

Several pilot areas discovered that the data included details of deceased people. Richmond reported that they had responses to letters indicating that just under 1% of the records returned from Royal Mail related to deceased people. Other areas also reported finding deceased individuals on the Royal Mail data.

The inclusion of deceased people could signify that the data is out-of-date, but equally, the redirection service may be used by family members who are redirecting the post of individuals who have died. Therefore the records relating to deceased individuals may well be current in terms of their use by Royal Mail, but are obviously not relevant to EROs. Royal Mail has confirmed that they could exclude deceased people from the data provided to EROs in future.

⁶⁵ Ref above – what are UPRNs

Existing electors

The proportion of individuals found to be existing electors is around 13% on average. This is slightly higher than the proportion found on the DWP home movers data (9%).

Ineligible individuals

The Royal Mail data requested by Cabinet Office included the names of individuals below the age for registering to vote. These are collected by Royal Mail as the mail redirection form asks for the details of all members of the household who are moving. We understand from Royal Mail that they can exclude these individuals from any data provided to EROs in future.

Some pilot areas realised that the data included individuals below the age for inclusion on the register while conducting checks on the data received. Other areas only became aware when they started receiving phone calls and enquiries in response to their letters.

Only a few areas reported the number of the ineligible individuals they found on the Royal Mail data, reporting they made up around 10-15% of the total sample.

Many of the pilot areas were unhappy about the fact that these records had been included with the data requested by Cabinet Office. One area chose not to do any more work with the Royal Mail data because of this issue. Some areas were also unhappy about the way Cabinet Office responded to this issue, noting that they were slow to inform them and were not particularly clear on what should be done.

Another issue with the Royal Mail data is that mail can be redirected to a correspondence address, such as a family member or a business premises, rather than an address that the individual currently lives at. Unlike the electoral register, there is no requirement that the individual actually resides at the address to which they have their mail forwarded.

In terms of ineligibility because of nationality, only three pilot areas reported figures and on average this is around 1%.

Other issues

Royal Mail was the only national database where there were significant issues in agreeing the terms and conditions of sharing data. It is clear that Royal Mail did not find many of the processes particularly straightforward, such as being required to go through a full tender process even though they are the only suppliers of this data.

While both Royal Mail and DWP charged for the pilot, the Royal Mail data cost around a third more than the DWP data, as Royal Mail charged for access as well as processing. (See Chapter 4 for details.)

There are restrictions on the data that Royal Mail can share. The data provided to EROs as part of this pilot were the details of individuals who had moved between addresses within that authority's boundaries as well as details of individuals who had moved in from outside.

Initially Cabinet Office had planned to include details of properties where the Royal Mail data indicated that the person listed on the register had moved out; however GDS did not have time to sort and return this data.

New registrations

The numbers of new registrations for individuals identified on the Royal Mail data ranged from 2% to 31%, as shown in Table D1 below.

Table D1: New registrations from Royal Mail data

| Area | Sample size ⁶⁶ | Number of letters | Number of new registrations | % new registrations (from sample) |
|----------------------|---|-------------------|-----------------------------|-----------------------------------|
| Ceredigion | 71 | 39 | 9 | 13% |
| Conwy | 786 | 682 | 173 | 22% |
| East Renfrewshire | 401 | 294 | 73 | 18% |
| Greenwich | 290 | 129 | 18 | 6% |
| Lothian | No follow up work conducted | | | |
| Pembrokeshire | Results not separated by database ⁶⁷ | | | |
| Powys | 157 | 153 | 48 | 31% |
| Richmond-upon-Thames | 3,416 | 2,842 | 210 | 6% |
| South Ribble | No follow up work conducted | | | |
| Southwark | 4,416 | 4,416 | 106 | 2% |
| Sunderland | 984 | 856 | 163 | 17% |
| Tower Hamlets | No follow up work conducted | | | |
| Wigan | Results not separated by database | | | |
| Wrexham | 74 | 61 | 5 | 7% |
| Total | 10,595 | 9,472 | 805 | 8% |

⁶⁶ See footnote 53 against Table A1 for details of how the samples have been calculated.

⁶⁷ Two databases were included for each option. Some pilot areas reported one set of figures for each option, combining results for individuals found on various databases.

For most pilot areas, the percentage of new registrations achieved from Royal Mail data is broadly similar to that achieved from DWP home movers data.

The only areas where there are notable differences are Conwy, Sunderland and Southwark. In Conwy the rate of new registrations on the Royal Mail data is twice as high as that on the DWP data. Conwy reported a lot of address issues in their DWP data, but not with their Royal Mail data, which may explain the difference in new registrations. The higher rate in Sunderland is largely explained by the records which were found on both databases having a much higher rate than records found on either DWP or Royal Mail – 33% compared to 8% for the DWP or Royal Mail records. The ‘found on both’ results are included in both the DWP and Royal Mail home mover results, but have a greater impact on the Royal Mail results because of the smaller sample size.

Conversely in Richmond the number of new registrations is two-thirds lower on the Royal Mail data than on the DWP data. Richmond had concerns about the quality of addressing on the Royal Mail data and so carried out less follow up work on this dataset.

The sample selected by East Renfrewshire is not a random sample, but is comprised of individuals who they had already contacted as part of their ongoing efforts to target home movers by using local house sales data. This is likely to explain why the number of new registrations is relatively high.

Wigan was unable to separate the two home movers databases in their report, but they achieved a 13% registration rate overall.

Pembrokeshire reported a 34% registration rate from letters although as set out in Appendix C we have been unable to clarify their approach to selecting which records to follow up.

Lothian VJB chose not to write-out to individuals identified on the Royal Mail data because of concerns over quality of the data. However they did conduct some checks against their register, and the results of these checks are included in our analysis where relevant. Tower Hamlets worked with the DWP data first and did not have time to do any follow up work with the Royal Mail data.

Conclusion

Overall, the Royal Mail data provided appears to have suffered from fewer of the addressing issues presented by other databases in the pilot, although the lack of UPRNs concerned a few pilot areas.

However, there was substantial variation in the number of new registrations achieved. In a couple of pilot areas the numbers of new registrations were quite high but overall, they were fairly low and broadly comparable with results for the DWP home movers data. These new registrations need to be balanced against the time involved in setting up the legal agreements and the high cost

of accessing this data, as well as the time spent by pilot areas in using the data.

The presence of individuals below the age for registering to vote was a concern to several pilot areas.

Any further data mining testing should only consider including Royal Mail data on the basis that individuals below the age of registering to vote will be filtered out, the currency will be restricted further and the start dates of mail redirections be included with the information provided to EROs.

Appendix E - Student Loans Company

This appendix sets out our findings in relation to the Student Loans Company (SLC) database.

Data held by the Student Loans Company data was used to try to identify students. The Government Digital Service (GDS) combined it with DfE and DfES data before sending the information to pilot areas, although as discussed in Chapter 5, both of these datasets are unsuitable for identifying unregistered students.

Quality of the data

Addressing

We only have usable results from one pilot area, Greenwich, for this database. They found that the address information was very incomplete, for example listing only a general area ('Woolwich, Greenwich') with no street information or postcode. Going through their file, Greenwich found that nearly a third of the SLC addresses were quite clearly incomplete.

SLC informed us that the addresses on the data they provided were complete: all had a postcode and street name and the majority had a house number. SLC were not aware of any reason why the addresses would be incomplete by the time they reached the pilot areas. It is unclear how or when this happened, but it seems that there may have been issues in the matching process.

There are two types of address held on the SLC database: the student's home or main address and their term-time address. Students do not have to enter a term-time address if they do not wish to, and they select one address only as their correspondence address.

The SLC conducts identity checks on new student finance applications and also checks ongoing attendance at the relevant institution every term. They do not check address as such, although they try to trace any individuals for whom correspondence is returned as undeliverable.

Both types of address were extracted from the SLC database and provided to GDS for inclusion in the pilot i.e. if a student had either a home or term-time address within one of the relevant pilot areas they were included in the dataset used for the matching (so, for example, they could be at university in Wolverhampton with a home address in Greenwich and be included in both extracts). Ideally we would have details of how many of each type of address

were returned to each pilot area, but unfortunately GDS are not able to provide this level of analysis.

Currency

Greenwich found that 12% of the records returned related to people who were no longer resident.

The timing of this pilot was not particularly suitable for targeting student registration, as the write-out took place near the end of the academic year, when a large number of students are likely to be leaving their term-time addresses. A better time to contact students would be October-November or January-February each year.

Existing electors

Greenwich found that 25% of the people on the SLC data were existing electors. This is likely to be largely due to the issues with addressing leading to failed matches. For example, Greenwich checked 170 incomplete SLC addresses against their register and found that 97% of the individuals were already registered.

Of the 26 records that Ceredigion received, 17 were existing electors.

Ineligible individuals

Greenwich reported finding that 5% of their sample were ineligible because they were ineligible foreign nationals. Student finance policy allows foreign nationals to receive student finance if certain criteria are met.

Other issues

Ceredigion received only 26 records from the SLC database, despite having a large student population. Cabinet Office were unable to explain why the number of records returned was so low. SLC told us that the data they provided was based on a list of postcodes received from Cabinet Office. It may be that this list was incomplete.

New registrations

We only have results from two areas in relation to the SLC data, since the other areas were unable to separate their results by database. In one of these areas, Ceredigion, the sample size is so small that the results are effectively unusable.

Table E1: New registrations from SLC data

| Area | Sample size⁶⁸ | Number of letters | Number of new registrations | % new registrations (from sample) |
|---------------|---|--------------------------|------------------------------------|--|
| Ceredigion | 26 | 4 | 0 | 0% |
| Coventry | Results not separated by database ⁶⁹ | | | |
| Greenwich | 764 | 425 | 32 | 4% |
| Tower Hamlets | Results not separated by database | | | |
| Wolverhampton | Results not separated by database | | | |
| Total | 790 | 429 | 32 | 4% |

Coventry and Wolverhampton provided results for students but were unable to separate databases in their reports. Coventry reported a 5% registration rate and Wolverhampton 15%. However, it is not possible to know which of these new registrations came from records identified on the SLC data, and which from records returned from DfE.⁷⁰

Conclusion

There seem to have been significant issues with the addresses in the SLC data received by the pilot areas. We are unable to say what caused these issues. These issues may explain why Greenwich was able to find so many of their sample as being already registered.

We have little quantitative data in relation to the records returned from the SLC database. However, the results that we do have indicate that there were few new registrations.

There may be merit in re-testing the SLC data, although only if Cabinet Office/GDS and SLC could resolve the addressing issues and if testing was conducted at a more suitable time of year.

⁶⁸ See footnote 53 against Table A1 for details of how the samples have been calculated.

⁶⁹ Two databases were included for each option. Some pilot areas reported one set of figures for each option, combining results for individuals found on various databases.

⁷⁰ For example, Wolverhampton received a single results file for 'students' which included data from SLC and DfE. As set out in Appendix X the DfE data did not include many, if any, students (in the sense of further or higher education). However, letters were still sent to the names on the results file and as responses came in no distinction was made between names held on one database or another, or both. Some of the registrations recorded by Wolverhampton as students may well therefore have been attainees (from the DfE file).

Appendix F - County data mining

This appendix sets out our findings on the four county council databases.

The main aim of this part of the pilot was to see if it was possible to establish data sharing arrangements between a county and district council for the purposes of electoral registration. This exercise was therefore more focused on developing a qualitative understanding of the barriers involved than in achieving new registrations.

Data sharing arrangements

Establishing data sharing arrangements between each pilot area and its respective county council took a significant amount of time and effort – this process was driven by Cabinet Office. Only one of the county councils responded to our request for an interview, so we do not know the specific issues that caused delays in each case.

However, it seems that at least some of the county councils had significant concerns about sharing data with their lower tier authority, and in some cases it was difficult to identify an individual or team within the county who would take responsibility for finalising the agreements.

In the end, it was possible to agree the terms and conditions of data sharing, and data was actually transferred, between all pairs of district and county councils. In this respect, the pilot achieved its aim of testing whether data sharing arrangements could be established.

However, some of the pilot areas commented that access seemed dependent on relationships with particular individuals at the county council, rather than on a more firmly embedded protocol.

It is difficult to draw firm conclusions from four pilot areas. However, difficulties in accessing data, even with the intervention and support of Cabinet Office, suggest that it would not necessarily be straightforward for other lower tier/district authorities to set up data sharing arrangements even if they had a legal right of access.

Matching process

Only one of the pilot areas was able to work with their EMS provider to establish an automated matching process. Another had hoped to do so, but ran out of time because of the delays.

Three pilot areas therefore matched the county data manually against their register. Two areas reported that this was a time-consuming process and took longer than expected, while the other was able to conduct the matching using UPRNs which made the process more efficient.

Quality and suitability of data

Each pilot area received a different dataset. As would be expected, the quality and suitability of the data they received varied.

One pilot area told us that the data seemed to be current and accurate, while another felt that it was very out-of-date, receiving responses to their letters informing them that the individual had moved out of the property many years previously.

One pilot area received education data covering the whole of the county rather than being restricted to their district. This caused them additional work in sorting the data before they could match it against their register. Cabinet Office do not believe there was any breach of data protection principles or of the pilot data sharing agreements as a result of the exchange of this wider dataset.

Of the three pilot areas who provided results, two reported finding no ineligible individuals on the county data, with the third finding only a very small number. This may reflect the demographics of these areas, with a smaller non-EU population than other parts of Great Britain.

Local data will be used during the transition to IER, particularly as part of the confirmation process. The variable quality of this local data reinforces the importance of each ERO taking time to satisfy themselves that any data they use is accurate.

New registrations

Originally, Cabinet Office did not envisage the pilot areas would conduct any additional follow up work with the county data, beyond matching it against their register. In the end, three areas did conduct some follow up work, but South Ribble did not contact any individuals during the pilot timescales.

In Rushmoor, the difference between the sample size and number of letters sent is due to the remaining addresses being unusable or clearly out of date.

Table F1: new registrations, county data mining

| Area | Sample size ⁷¹ | Number of letters sent | New registrations | % new registrations (from sample) |
|-------------------|---------------------------|------------------------|-------------------|-----------------------------------|
| Barrow-in-Furness | 346 | 346 | 91 | 26% |
| Mansfield | 628 | 628 | 56 | 9% |
| Rushmoor | 853 | 765 | 75 | 9% |
| Total | 1,827 | 1,739 | 222 | 12% |

Conclusion

This pilot has demonstrated that it is possible to establish data sharing arrangements between district and county councils, albeit for a one-off transfer and with a great deal of effort.

All four of the pilot areas were able to match the county data against their register, although in some cases this took far longer than they had expected.

Three areas wrote to the individuals they could not find on their register. Response rates to these letters seem reasonable: on the higher end of the response rates achieved using national data. However, the sample sizes are small and we have only three sets of results, so care should be taken in relying on this data.

Overall, experiences of using the county data were mixed and we are unable to conclude whether access to county data would assist EROs in district authorities in maintaining their electoral registers.

⁷¹ Sample size is the number of unregistered individuals the area found on the county data, excluding any individuals found too young to register, in line with the approach we have adopted for national data mining results.

Appendix G – National data mining, approaches to follow up work

| Area | Target group(s) | Sampling | Checks | Write-out | Canvassing |
|------------|--------------------------------------|---|--|---|---------------|
| Ceredigion | Attainers Home movers Students | Excluded records with address issues e.g. missing address lines; otherwise followed up all records received | All records in sample checked against register Checks conducted manually | 936 registration forms sent out (across all three options) on 8 May | No canvassing |
| Conwy | Home movers | Conwy followed up all records received | All records checked against register Checks conducted manually | 3,054 registration forms sent out on 2 May | No canvassing |
| Coventry | Attainers Students | Followed up all records received except for individuals who were under 16 (485 records on the attainers file) and individuals who were already 18 (2,881 records on the attainers file). The latter were sent registration forms separately to the pilot. | Some records in sample checked against register Checks were conducted by Coventry's IT department and staff in Electoral Services | Coventry sent out 11,475 registration forms at the end of April One reminder sent to non-responders in mid-May | No canvassing |

| Area | Target group(s) | Sampling | Checks | Write-out | Canvassing |
|-------------------|-------------------------|---|---|--|---|
| East Renfrewshire | Home movers | All DWP records with UPRNs (2,469 records) 401 Royal Mail records followed up; sample selected as Renfrewshire VJB already had intelligence from local data that these individuals had moved house and had contacted them as part of their ongoing registration efforts. | All DWP sample checked against register. Address matched automatically against register, with manual checks of names. Royal Mail sample – all checked against register manually. | Approx. 20% of the DWP sample (535 records) sent registration forms between 8-10 May Registration forms sent to 294 Royal Mail records between 8-10 May | Non-responders canvassed once - DWP sample only |
| Greenwich | Home movers Students | Greenwich selected a random sample of approximately 14% of records received | All records in sample checked against register Checks conducted manually | 2,491 registration forms sent out in batches between 22 April and 7 May | No canvassing |
| Harrow | Attainers | Followed up records found on Department for Education only (excluding records found on DWP and records found on both databases). This was due to poor communication from Cabinet Office and incorrect labels on the data returned. | All records in sample checked against register Checks conducted manually | 519 letters sent out in late April | No canvassing |

| Area | Target group(s) | Sampling | Checks | Write-out | Canvassing |
|---------------|---------------------------|---|---|---|---------------|
| Lothian | Home movers | All records within the postcodes originally specified | All records in sample checked against register Checks conducted manually | Lothian VJB sent out 373 registration forms on 6 May, to DWP records only Royal Mail records not contacted because of concerns over data | No canvassing |
| Pembrokeshire | Home movers ⁷² | The sample excludes all records with address issues, including records with missing address information or references to defunct county names The total sample size is unclear as we only have figures for the number of registration forms sent out, but Pembrokeshire found an unknown number of records related to existing electors before sending forms | All records in sample checked against register Checks conducted manually | 540 registration forms sent out in late April | No canvassing |

⁷² Pembrokeshire were due to participate in the attainers testing as well but did not receive this data - see footnote 55

| Area | Target group(s) | Sampling | Checks | Write-out | Canvassing |
|-------------|--------------------------|--|---|---|---|
| Powys | Attainers Home movers | <p>Powys excluded records outside of the wards and postcodes they had originally specified (pilot areas were invited to specify particular wards or postcodes to focus on but in the end it was not possible for GDS to restrict all of the data returned to these areas).</p> <p>It is possible that Powys also excluded records with address issues at this stage.</p> | <p>Checked all records in sample against register</p> <p>Checks conducted manually</p> | <p>1,301 registration forms sent on 1 May</p> <p>784 reminder letters sent to non-responders on 17 May</p> | <p>Canvassed non-responders within one electoral division only on 21-22 May</p> |
| Richmond | Home movers | All records received | <p>All records checked against register</p> <p>Combination of automated and manual checks</p> | <p>Richmond sent out 3,317 registration forms on 19 April</p> <p>Reminder letter sent to DWP records and RM within specified postcodes only (744 letters)</p> | <p>Canvassed DWP non-responders</p> |

| Area | Target group(s) | Sampling | Checks | Write-out | Canvassing |
|--------------|--------------------------|--|--|--|---------------|
| Rushmoor | Attainers | Excluded 1,319 records as individuals were too old or too young to be registered as attainers. | All records in sample checked against register Checks conducted manually | Registration forms sent out on 7 May (county data) and 17 May (DWP data) Write-out included flier tailored towards this age group and prize draw incentive for returning form | No canvassing |
| South Ribble | Home movers | <i>No follow up during timescales of pilot</i> | | | |
| Southwark | Home movers | All records followed up | No checks before registration forms sent Completed forms checked against register and council tax records | 9,077 registration forms sent out on 8 May | No canvassing |
| Sunderland | Attainers Home movers | Followed up all records received | Checked all records against the register, council tax and housing benefit databases Checks conducted by automated processes | Sunderland sent out 19,351 registration forms on 20 May | No canvassing |

| Area | Target group(s) | Sampling | Checks | Write-out | Canvassing |
|---------------|--------------------------------------|--|---|---|---|
| Tower Hamlets | Attainers Home movers Students | All files – records within 7 wards (the wards with lowest canvass response rate in 2012) Home movers file – all DWP records updated within the past three months and some DWP records updated within the past six months | All records in sample checked against register and other council datasets Checks conducted manually | Tower Hamlets sent out registration forms between 15-31 May | No canvassing |
| Wigan | Home movers | Wigan followed up all records received | Royal Mail records checked against register Checks conducted manually | 13,502 registration forms sent out in early May | No canvassing |
| Wolverhampton | Attainers Students | Wolverhampton followed up all records received, except for duplicate records. There were a large number of duplicate records between the attainers and students file, because the DfE data was identical on each. Duplicate records were excluded from follow up as they were identified. | Records with addresses in certain wards (15 out of 20) were checked against the register, with some checks against council tax records These checks were mostly conducted manually | Records in 17 wards were sent a registration form between 19 April and 14 May | Records in three wards were canvassed between 26 April and 6 May (rather than receiving a form by post), with canvassers delivering the form if no-one was in |

| Area | Target group(s) | Sampling | Checks | Write-out | Canvassing |
|-------------|------------------------|----------------------------------|---|---|-------------------|
| Wrexham | Home movers | Followed-up all records received | Checked all records in sample against register Checks conducted manually | Wrexham sent out 2,096 registration forms at the end of April | No canvassing |

Appendix H – Pilot area profiles

This appendix contains brief profiles of the 20 areas that took part in the pilot.

- | | |
|----------------------------------|-------------------|
| 1. Barrow-in-Furness | 11. Powys |
| 2. Ceredigion | 12. Richmond |
| 3. Conwy | 13. Rushmoor |
| 4. Coventry | 14. South Ribble |
| 5. East Renfrewshire | 15. Southwark |
| 6. Greenwich | 16. Sunderland |
| 7. Harrow | 17. Tower Hamlets |
| 8. Lothian Valuation Joint Board | 18. Wigan |
| 9. Mansfield | 19. Wolverhampton |
| 10. Pembrokeshire | 20. Wrexham |

Sources

Demographic information is drawn from the 2011 census (available from the Office of National Statistics (ONS)), other than the figure for estimated population mobility, which is taken from the 2001 census.

Figures for the size of the electoral register and the number of attainers on the register for each area are taken from the official electoral statistics 2012 (ONS). Other electoral registration data is taken from data submitted by Electoral Registration Officers to the Electoral Commission each year. This data is available on our website:

<http://www.electoralcommission.org.uk/performance-standards/electoral-registration>

The national databases referred to on the profiles are:

- Department for Education (DfE)
- Department for Work and Pensions (DWP)
- Royal Mail
- Student Loans Company (SLC)

Barrow-in-Furness Borough Council

Demographic and electoral registration information

| Demographic information (2011 census data) | | Rank ⁷³ |
|---|--------|--------------------|
| Population (16+) | 56,796 | - |
| Attainers (17 year olds) | 986 | - |
| Full-time students | 2.0% | 319 th |
| Private renters | 13.4% | 198 th |
| Estimated population mobility ⁷⁴ | 12.8% | 345 th |

Electoral registration data

| | |
|---|--------|
| Local government register entries (October 2012) | 53,956 |
| Number of attainers on register (October 2012) | 591 |
| Canvass return rate (average 2010-2012) | 89.5% |
| Proportion of electors added in canvass (average 2010-2012) | 8.5% |

Results

Matching process

| Database | Number of records received | Number excluded from follow-up | Number found on register | Number not found on register (potential new electors) |
|---------------------------------------|----------------------------|--------------------------------|--------------------------|---|
| Cumbria County Council education data | 10,459 | 9,711 | 18 | 730 |

⁷³ Relative to the 348 authorities in England and Wales (410 for 2001); highest to lowest proportions.

⁷⁴ This figure is taken from the 2001 census data as the 2011 figures are not yet available.

Follow-up work

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|---------------------------------------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| Cumbria County Council education data | 346 | 346 | 91 | 26% |

Summary

Barrow excluded 9,712 records from follow-up because they were outside their district. Cumbria provided education records covering the whole of the county, although Barrow had not been expecting this. Cabinet Office do not believe there was any breach of data protection or of the pilot data sharing agreements as a result of the exchange of this wider dataset.

Of the 748 records that were within Barrow, they were able to match 18 (2%) to existing electors on their register.

Barrow worked with their Electoral Management Software supplier to develop an automated matching process.

346 of the 730 potential new electors identified were followed up, as the remaining 384 records related to individuals who were under 16.

The number of new registrations is high compared to other pilot areas.

Barrow are supportive of the principle of using local and county data to identify potential new electors and have recently started receiving council tax updates which they will use to target home movers.

Ceredigion County Council

Demographic and electoral registration information

| Demographic information (2011 census data) | | Rank ⁷⁵ |
|---|--------|--------------------|
| Population (16+) | 64,692 | - |
| Attainers (17 year olds) | 798 | - |
| Full-time students | 15.2% | 6 th |
| Private renters | 21.0% | 60 th |
| Estimated population mobility ⁷⁶ | 22.2% | 22 nd |

Electoral registration data

| | |
|---|--------|
| Local government register entries (October 2012) | 58,320 |
| Number of attainers on register (October 2012) | 328 |
| Canvass return rate (average 2010-2012) ⁷⁷ | 78.35% |
| Proportion of electors added in canvass (average 2010-2012) | 16.8% |

Results

Attainers

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|----------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DWP | 1,763 | 611 | 87 | 5% |

⁷⁵ Relative to the 348 authorities in England and Wales (410 for 2001); highest to lowest proportions.

⁷⁶ This figure is taken from the 2001 census data as the 2011 figures are not yet available.

⁷⁷ Data for the canvass return rate and proportion of electors added is only available for 2010 and 2012

Home movers

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|--------------------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DWP | 330 | 282 | 29 | 9% |
| Royal Mail | 66 | 34 | 8 | 12% |
| Both ⁷⁸ | 5 | 5 | 1 | 20% |

Students

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|----------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| SLC | 26 | 4 | 0 | 0% |

Summary

Ceredigion also received data from the Welsh Government, Department for Education and Skills database (on the attainers and students files). In line with other pilot areas who received this data, they did no follow-up work with these records as they did not contain full addresses.

Ceredigion received only 26 records from the SLC database, despite having a large student population. Cabinet Office were unable to explain why the number of records returned was so low. SLC told us that the data they provided was based on a list of postcodes received from Cabinet Office, so it may be that this list was incomplete.

Ceredigion found that 13% of the home movers records related to existing electors. Although they received only 26 records from SLC, they found 17 of these were existing electors.

Overall, Ceredigion felt that the amount of work involved in using this data was high in light of the overall number of new registrations.

⁷⁸ This category refers to individuals whose details were found on both databases used for that option e.g. found on both the DWP and Royal Mail databases for home movers.

Conwy County Borough Council

Demographic and electoral registration information

| Demographic information (2011 census data) | | Rank ⁷⁹ |
|---|--------|--------------------|
| Population (16+) | 96,102 | - |
| Attainers (17 year olds) | 1,441 | - |
| Full-time students | 3.0% | 177 th |
| Private renters | 16.1% | 115 th |
| Estimated population mobility ⁸⁰ | 14.4% | 234 th |

Electoral registration data

| | |
|---|--------|
| Local government register entries (October 2012) | 93,164 |
| Number of attainers on register (October 2012) | 856 |
| Canvass return rate (average 2010-2012) | 97.7% |
| Proportion of electors added in canvass (average 2010-2012) ⁸¹ | 10.3% |

Results

Home movers

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|--------------------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DWP | 3,072 | 2,372 | 333 | 11% |
| Royal Mail | 714 | 616 | 149 | 21% |
| Both ⁸² | 72 | 66 | 24 | 33% |

⁷⁹ Relative to the 348 authorities in England and Wales (410 for 2001); highest to lowest proportions.

⁸⁰ This figure is taken from the 2001 census data as the 2011 figures are not yet available.

⁸¹ Data only available for 2011 and 2012

⁸² This category refers to individuals whose details were found on both databases used for that option e.g. found on both the DWP and Royal Mail databases for home movers.

Summary

Conwy discovered a lot of problems with the addresses on the DWP data, including flat and house numbers, and even street names, missing. Many of the addresses also used the previous county names, which have not existed since 1996.

Conwy found that 23% of the DWP records referred to existing electors. The figure for the Royal Mail data was only 14%. The problems with the DWP addresses are likely to explain at least some of this, as poor quality addresses mean that existing electors are less likely to be excluded from the data returned to EROs.

Conwy did not identify as many address issues with the Royal Mail data and this is likely to account for the higher rate of new registrations for this dataset.

Individuals found on both the DWP and Royal Mail data have the highest rate of new registrations. This may indicate that records returned from two national data sources are more likely to be current and accurate. However, the numbers involved are small and care should be taken in interpreting these results.

Coventry City Council

Demographic and electoral registration information

| Demographic information (2011 census data) | | Rank ⁸³ |
|---|---------|--------------------|
| Population (16+) | 254,106 | - |
| Attainers (17 year olds) | 3,883 | - |
| Full-time students | 11.6% | 22 nd |
| Private renters | 21.5% | 54 th |
| Estimated population mobility ⁸⁴ | 17.1% | 86 th |

Electoral registration data

| | |
|---|---------|
| Local government register entries (October 2012) | 237,472 |
| Number of attainers on register (October 2012) | 2,481 |
| Canvass return rate (average 2010-2012) | 94.3% |
| Proportion of electors added in canvass (average 2010-2012) | 15.7% |

Results

Attainers

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|------------------------------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DfE and/or DWP ⁸⁵ | 5,361 | 5,280 | 588 | 11% |

⁸³ Relative to the 348 authorities in England and Wales (410 for 2001); highest to lowest proportions.

⁸⁴ This figure is taken from the 2001 census data as the 2011 figures are not yet available.

⁸⁵ Coventry did not report their results separately by database.

Students

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|----------------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DfE and/or SLC | 6,939 | 6,195 | 337 | 5% |

Summary

Coventry believed that this data could be useful to EROs, although found that they needed to conduct a lot of checks and data cleansing before sending out registration forms. They do not currently have the resources or skills within the electoral services team to do automated checks, so their IT department assisted with this pilot.

The number of new registrations is much lower on the students file than the attainers file. Coventry felt that generally they did not get a very high response rate and were targeting a group of electors (young people) who are not particularly engaged with politics.

The difference in the registration rates between the two files may be due to problems with SLC data. Although Coventry did not report their results separately by database, we know from other pilot areas that there were significant problems with the addresses on the SLC data, meaning that many existing electors were returned on the data and that addresses were unusable for a write-out. Coventry found that 8% of the records on the students file were already on the register, compared with 1% of the attainers file.

East Renfrewshire

Renfrewshire Valuation Joint Board (covering East Renfrewshire, Inverclyde and Renfrewshire) participated in the pilot, but using only East Renfrewshire's electoral register.

Demographic and electoral registration information

Demographic information (2011 census data)

| | |
|------------------|--------|
| Population (16+) | 73,900 |
|------------------|--------|

The remaining 2011 census data is not yet available for Scotland.

Electoral registration data

| | |
|---|--------|
| Local government register entries (December 2012) | 69,447 |
| Number of attainers on register (December 2012) | 1,097 |
| Canvass return rate (average 2010-2012) | 96.5% |
| Proportion of electors added in canvass (average 2010-2012) | 6.7% |

Results

Home movers

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|------------|-------------------|-----------------------------------|-----------------------------|-----------------------------------|
| DWP | 610 ⁸⁶ | 535 | 59 | 10% |
| Royal Mail | 401 | 294 | 73 | 18% |

⁸⁶ Renfrewshire VJB conducted checks against the register for all DWP records with UPRNs in the East Renfrewshire local authority area (2,469 records). They found 305 on the register (12%). They then took a random sample of around 24% of the records not on the register to write to. We have added back in the proportion of records they found on the register to their sample in order to provide comparable figures.

Summary

On the Royal Mail data, the sample selected by East Renfrewshire is not a random sample, but is comprised of individuals who they had already contacted as part of their ongoing efforts to target home movers by using local house sales data. This is likely to at least partly explain why the number of new registrations is higher on the Royal Mail than DWP data.

Renfrewshire VJB found that around 12-14% of the DWP records related to existing electors. A comparable figure is not available for the Royal Mail records, as they specifically targeted individuals who they had already invited to register.

Renfrewshire VJB felt that generally the results (in terms of new registrations) from the national data mining were not as good as the results they achieve using local data to target home movers, where they estimate around 25-35% of the records received result in new registrations.

Royal Borough of Greenwich

Demographic and electoral registration information

| Demographic information (2011 census data) | | Rank ⁸⁷ |
|---|---------|--------------------|
| Population (16+) | 199,163 | - |
| Attainers (17 year olds) | 3,099 | - |
| Full-time students | 8.8% | 45 th |
| Private renters | 21.3% | 56 th |
| Estimated population mobility ⁸⁸ | 18.0% | 68 th |

Electoral registration data

| | |
|---|---------|
| Local government register entries (December 2012) | 173,421 |
| Number of attainers on register (December 2012) | 1,793 |
| Canvass return rate (average 2010-2012) | 93.4% |
| Proportion of electors added in canvass (average 2010-2012) | 11.6% |

Results

Home movers

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|------------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DWP | 3,148 | 1,623 | 132 | 4% |
| Royal Mail | 290 | 129 | 18 | 6% |

⁸⁷ Relative to the 348 authorities in England and Wales (410 for 2001); highest to lowest proportions.

⁸⁸ This figure is taken from the 2001 census data as the 2011 figures are not yet available.

Students

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|----------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DfE | 430 | 314 | 41 | 10% |
| SLC | 764 | 425 | 32 | 4% |

Summary

Greenwich conducted extensive checks on the records in their samples before sending out registration forms. Just over half of the records received a registration form. The remaining records were excluded as being out-of-date (Greenwich could see that the individual had already moved out of that property), relating to existing electors or having incomplete address information.

On the home movers option, Greenwich found that 26% of the DWP records and 16% of the Royal Mail records related to existing electors. Greenwich reported poor quality address information on the DWP data and some duplicate records on the Royal Mail data.

On the students option, Greenwich reported finding extensive poor quality or incomplete addresses on the SLC data, estimating that approximately 28% of addresses were incomplete. They found that 25% of the SLC sample related to existing electors. The DfE data had the fewest address issues and had the highest rate of new registrations – 10% of the sample. However Greenwich were still able to find 18% of the sample were already registered.

London Borough of Harrow

Demographic and electoral registration information

| Demographic information (2011 census data) | | Rank ⁸⁹ |
|---|---------|--------------------|
| Population (16+) | 190,998 | - |
| Attainers (17 year olds) | 3,451 | - |
| Full-time students | 6.8% | 69 th |
| Private renters | 24.0% | 37 th |
| Estimated population mobility ⁹⁰ | 16.3% | 117 th |

Electoral registration data

| | |
|---|---------|
| Local government register entries (December 2012) | 177,953 |
| Number of attainers on register (December 2012) | 1,858 |
| Canvass return rate (average 2010-2012) | 91.8% |
| Proportion of electors added in canvass (average 2010-2012) | 10.6% |

Results

Attainers

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|--------------------------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| Department for Education | 584 | 519 | 13 | 2% |

⁸⁹ Relative to the 348 authorities in England and Wales (410 for 2001); highest to lowest proportions.

⁹⁰ This figure is taken from the 2001 census data as the 2011 figures are not yet available.

Summary

Poor communication from Cabinet Office and incorrect data labels meant that Harrow understood the file they received to contain data from the DfE database only, with records that both could and could not be matched against their register. They had actually received records from both databases but only those that could not be matched against the register. However, the DWP data was labelled 'found on Harrow' and they were also informed by Cabinet Office that there had been a problem with the DWP data.

Harrow checked a random sample of the records in the 'found on Harrow' category against their register and found that a high number of these were existing electors (this is in line with findings from other pilot areas which indicate that the data returned did include significant numbers of existing electors, even those these should largely have been excluded). Harrow were therefore satisfied that this data had been matched centrally and that there would be no benefit in them conducting any further checks or follow up work.

Several pilot areas raised concerns about the support and guidance provided by Cabinet Office. Harrow is not alone in experiencing problems which could have been avoided if there had been clearer and more regular communication.

The fact that Harrow followed up a small number of records which were found only on one database is likely to account, at least in part, for the low numbers of new registrations. Harrow also felt that late April to May is an inappropriate time for contacting attainers, as many of them will be sitting exams at this time.

Lothian Valuation Joint Board

Covering East Lothian, the City of Edinburgh, Midlothian and West Lothian

Demographic and electoral registration information

Demographic information (2011 census data)

| | |
|------------------|---------|
| Population (15+) | 701,700 |
|------------------|---------|

The remaining 2011 census data is not yet available for Scotland.

Electoral registration data

| | |
|---|---------|
| Local government register entries (December 2012) | 619,162 |
| Number of attainers on register (December 2012) | 5,719 |
| Canvass return rate (average 2010-2012) | 86% |
| Proportion of electors added in canvass (average 2010-2012) | 9% |

Results

Home movers

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|----------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DWP | 453 | 373 | 30 | 7% |

Lothian VJB did not contact any individuals found on the Royal Mail data as they had concerns about the data.

Summary

Lothian VJB received a high volume of data back from the central matching process (around 14,000 records in total), in part because the Royal Mail data was not limited to the postcodes they had specified. Their IT department wrote a program to extract the records that related to the original postcodes, to reduce the number of records to a manageable level.

Lothian VJB found the address information on the records received difficult to work with. The addresses generally did not have UPRNs and were in a different format to the addresses held on the registers.

During the checks, Lothian VJB realised that details of individuals below the age of registering to vote were included on the Royal Mail data. They therefore decided to do no further work with this data and were unhappy that they had received information that was unsuitable for the purposes of electoral registration.

Lothian found that 13-14% of the records returned related to existing electors (both DWP and Royal Mail data).

Lothian do not feel that data mining in its current form is useful for EROs, as the address information is not particularly important on the national databases but it is for electoral registration.

Mansfield District Council

Demographic and electoral registration information

| Demographic information | | Rank ⁹¹ |
|---|--------|--------------------|
| Population (16+) | 85,448 | - |
| Attainers (17 year olds) | 1,402 | - |
| Full-time students | 2.4% | 258 th |
| Private renters | 14.1% | 178 th |
| Estimated population mobility ⁹² | 13.5% | 309 th |

Electoral registration data

| | |
|---|--------|
| Local government register entries (October 2012) | 82,180 |
| Number of attainers on register (October 2012) | 1,054 |
| Canvass return rate (average 2010-2012) | 92.2% |
| Proportion of electors added in canvass (average 2010-2012) | 9.2% |

Results

Matching process

| Database | Number of records received | Number excluded from follow-up | Number found on register | Number not found on register (potential new electors) |
|---|----------------------------|--------------------------------|--------------------------|---|
| Nottinghamshire County Council children's services data | 1,450 | - | 822 | 628 |

⁹¹ Relative to the 348 authorities in England and Wales (410 for 2001); highest to lowest proportions.

⁹² This figure is taken from the 2001 census data as the 2011 figures are not yet available.

Follow-up work

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|---|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| Nottinghamshire County Council children's services data | 628 | 628 | 56 | 9% |

Summary

Mansfield manually matched the county data against their register and found this to be quite time-consuming. The data received needed extensive re-formatting before they could easily match it against their register.

Mansfield were able to match 822 (57%) of the records receiving to records on their register.

The number of new registrations is broadly in line with the pilot average.

Mansfield would have liked more time to conduct follow-up work and report on outcomes, especially as the receipt of data coincided with the elections period meaning that they could not start work on it straight away.

Pembrokeshire County Council

Demographic and electoral registration information

| Demographic information (2011 census data) | | Rank ⁹³ |
|---|---------|--------------------|
| Population (16+) | 100,362 | - |
| Attainers (17 year olds) | 1,638 | - |
| Full-time students | 2.3% | 292 nd |
| Private renters | 13.1% | 212 th |
| Estimated population mobility ⁹⁴ | 14.6% | 221 st |

Electoral registration data

| | |
|---|--------|
| Local government register entries (October 2012) | 93,147 |
| Number of attainers on register (October 2012) | 667 |
| Canvass return rate (average 2010-2012) ⁹⁵ | 94.5% |
| Proportion of electors added in canvass (average 2010-2012) | 9.8% |

Results

Home movers

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|-------------------------------------|---------------|-----------------------------------|-----------------------------|-----------------------------------|
| DWP and/or Royal Mail ⁹⁶ | Not available | 540 | 181 | - |

⁹³ Relative to the 348 authorities in England and Wales (410 for 2001); highest to lowest proportions.

⁹⁴ This figure is taken from the 2001 census data as the 2011 figures are not yet available.

⁹⁵ Figures for the canvass return rate and proportion of electors added are only available for 2010 and 2012

⁹⁶ Pembrokeshire did not report their results separately by database.

Summary

Pembrokeshire discovered various problems with the addresses on the home movers data they received, including the addresses missing crucial pieces of information and containing references to counties which had not existed since 1996.

Pembrokeshire excluded all records with address issues from follow-up and then took a random sample from the remainder. We are not clear how many records were in this sample as we do not know how many individuals they were able to find on their register before sending out registration forms.

Pembrokeshire did not split their results by database, but it may be that by excluding all records with address issues, they in fact only followed up Royal Mail records, as we know from other pilot areas that DWP records seemed to have far more extensive address issues.

From the registration forms sent out, Pembrokeshire achieved a new registration rate of 34%, one of the highest in the pilot. However, they felt strongly that there was far too much work involved in sorting and cleansing the data.

Powys County Council

Demographic and electoral registration information

| Demographic information (2011 census data) | | Rank ⁹⁷ |
|---|---------|--------------------|
| Population (16+) | 110,083 | - |
| Attainers (17 year olds) | 1,789 | - |
| Full-time students | 2.1% | 318 th |
| Private renters | 14.3% | 171 st |
| Estimated population mobility ⁹⁸ | 13.6% | 300 th |

Electoral registration data

| | |
|---|---------|
| Local government register entries (October 2012) | 103,518 |
| Number of attainers on register (October 2012) | 1,152 |
| Canvass return rate (average 2010-2012) ⁹⁹ | 93.7% |
| Proportion of electors added in canvass (average 2010-2012) | 6.9% |

Results

Attainers

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|----------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DWP | 391 | 377 | 120 | 31% |

⁹⁷ Relative to the 348 authorities in England and Wales (410 for 2001); highest to lowest proportions.

⁹⁸ This figure is taken from the 2001 census data as the 2011 figures are not yet available.

⁹⁹ Figures for the canvass return rate and proportion of electors added are only available for 2010 and 2012

Home movers

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|------------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DWP | 823 | 771 | 223 | 27% |
| Royal Mail | 110 | 110 | 21 | 19% |
| Both | 47 | 43 | 27 | 57% |

Powys did no follow-up work with records returned from the Welsh Department for Education and Skills database as these did not contain full addresses.

Summary

Powys achieved some of the highest rates of new registrations in the pilot, whether measured as a proportion of the sample they worked with or as a proportion of registration forms sent out. This may be due to the fact that they carried out more follow-up work than other areas, sending a reminder letter and conducting some limited canvassing. The higher response rates may also reflect the demographics of this area.

Powys found that 15% of the DWP sample and 23% of the Royal Mail sample were no longer resident or not known at address. This perhaps indicates the data on the national databases being out-of-date, and the need to set tighter currency restrictions, even in an area like Powys with a relatively low level of population mobility.

Powys felt that data mining could potentially be a useful way to identify new electors, with the DWP data proving most effective overall.

London Borough of Richmond-upon-Thames

Demographic and electoral registration information

| Demographic information (2011 census data) | | Rank ¹⁰⁰ |
|--|---------|---------------------|
| Population (16+) | 150,052 | - |
| Attainers (17 year olds) | 1,821 | - |
| Full-time students | 4.9% | 89 th |
| Private renters | 21.1% | 58 th |
| Estimated population mobility ¹⁰¹ | 21.1% | 31 st |

Electoral registration data

| | |
|---|---------|
| Local government register entries (December 2012) | 138,843 |
| Number of attainers on register (December 2012) | 1,217 |
| Canvass return rate (average 2010-2012) | 97.3% |
| Proportion of electors added in canvass (average 2010-2012) | 12.0% |

Results

Home movers

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|------------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DWP | 609 | 475 | 102 | 17% |
| Royal Mail | 3,416 | 2,842 | 210 | 6% |

¹⁰⁰ Relative to the 348 authorities in England and Wales (410 for 2001); highest to lowest proportions.

¹⁰¹ This figure is taken from the 2001 census data as the 2011 figures are not yet available.

Summary

Richmond was the only area looking at the home movers option where the number of new registrations from the DWP data was significantly higher than on the Royal Mail data. Richmond carried out more follow-up work with the DWP data as they had serious concerns about the quality and suitability of the Royal Mail data. Issues they reported with the Royal Mail data include poor quality addressing and inclusion of business addresses.

Richmond was one of only a few pilot areas who conducted any canvassing. They canvassed only DWP non-responders for the reasons outlined above. Nearly a quarter of their new registrations came from the canvassing, demonstrating the impact this can have on registration rates.

Although the registration rate from the DWP data was good, Richmond still found that 22% of the DWP records related to individuals who were no longer resident or not known at address. This compares to 10% on the Royal Mail data, although the lower rate here may be due to fewer records being followed up and so results being less complete. Richmond found that 5% of records on both files related to existing electors.

Rushmoor Borough Council

Demographic and electoral registration information

| Demographic information (2011 census data) | | Rank ¹⁰² |
|--|--------|---------------------|
| Population (16+) | 74,854 | - |
| Attainers (17 year olds) | 1,233 | - |
| Full-time students | 3.3% | 141 st |
| Private renters | 19.2% | 69 th |
| Estimated population mobility ¹⁰³ | 20.2% | 38 th |

Electoral registration data

| | |
|---|--------|
| Local government register entries (October 2012) | 66,328 |
| Number of attainers on register (October 2012) | 645 |
| Canvass return rate (average 2010-2012) | 96.1% |
| Proportion of electors added in canvass (average 2010-2012) | 9.2% |

Results – national data mining

Attainers

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|--------------------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DWP ¹⁰⁴ | 1,771 | 1,301 | 166 | 9% |

¹⁰² Relative to the 348 authorities in England and Wales (410 for 2001); highest to lowest proportions.

¹⁰³ This figure is taken from the 2001 census data as the 2011 figures are not yet available.

¹⁰⁴ Rushmoor did not separate their results from database. According to GDS, they only received 14 records solely from DfE (the second database used for identifying attainers). Their results are therefore reported as DWP results, as being more representative of this database.

Results – county data mining

Matching process

| Database | Number of records received | Number excluded from follow-up | Number found on register | Number not found on register (potential new electors) |
|---|----------------------------|--------------------------------|--------------------------|---|
| Hampshire County Council education data | 2,528 | 1,210 | 465 | 853 |

Follow-up work

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|---|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| Hampshire County Council education data | 853 | 765 | 75 | 9% |

Summary

Rushmoor were the only pilot area who participated in both the national and county data mining. They targeted unregistered attainers in both. Nearly half of the individuals identified from the mining were identified on both the county and DWP data. These individuals are included in both sets of figures above, meaning that there is some double counting. Overall, Rushmoor worked with 2,118 records and in the timescales of the pilot, completed 171 new registrations (8%). Since submitting their results, Rushmoor have completed an additional 40 new registrations, which raises their overall registration rate to 10%.

Rushmoor had concerns about the currency and address information on both the county and DWP data. They found that on 10% of the county records the individual had already moved out of the property or the address was incomplete. On the DWP data, only 2% of the records were found to relate to individuals no longer at that address; however, this category includes

individuals where Rushmoor was informed that they had left the address seven or more years previously.

The extract sent to Cabinet Office for matching did not include any of the 16 year olds Rushmoor had details for (i.e. individuals who were not quite attainers at the date of the canvass but whose information had been included on the canvass form). This is likely to at least partly explain the high proportion of DWP records they found to be existing electors, around 24%.

This issue also means that the size of the DWP sample is likely to be inflated. It is likely that the proportion of new registrations reported should be higher, although we cannot say by how much.

South Ribble Borough Council

Demographic and electoral registration information

| Demographic information (2011 census data) | | Rank ¹⁰⁵ |
|--|--------|---------------------|
| Population (16+) | 89,163 | - |
| Attainers (17 year olds) | 1,422 | - |
| Full-time students | 2.9% | 185 th |
| Private renters | 9.2% | 335 th |
| Estimated population mobility ¹⁰⁶ | 11.8% | 383 rd |

Electoral registration data

| | |
|---|--------|
| Local government register entries (October 2012) | 86,347 |
| Number of attainers on register (October 2012) | 927 |
| Canvass return rate (average 2010-2012) ¹⁰⁷ | 96.2% |
| Proportion of electors added in canvass (average 2010-2012) | 7.0% |

Results

South Ribble participated in the county and home movers data mining.

For the county mining, South Ribble received data from Lancashire County Council in late April and matched this against their register, using manual checks. They found approximately one-third of the individuals identified on the county data were already on their register, leaving approximately 200 potential new electors. South Ribble did not contact any of the potential electors during the timescales of the pilot, partly because of the overlap with the elections period.

South Ribble received home movers data in mid-April but again did not contact any of the individuals identified during the timescales of the pilot.

¹⁰⁵ Relative to the 348 authorities in England and Wales (410 for 2001); highest to lowest proportions.

¹⁰⁶ This figure is taken from the 2001 census data as the 2011 figures are not yet available.

¹⁰⁷ Data for the canvass return and proportion of electors added is only available for 2011 and 2012

London Borough of Southwark

Demographic and electoral registration information

| Demographic information (2011 census data) | | Rank ¹⁰⁸ |
|--|---------|---------------------|
| Population (16+) | 234,901 | - |
| Attainers (17 year olds) | 2,830 | - |
| Full-time students | 11.5% | 24 th |
| Private renters | 25.0% | 35 th |
| Estimated population mobility ¹⁰⁹ | 22.3% | 21 st |

Electoral registration data

| | |
|---|---------|
| Local government register entries (December 2012) | 203,514 |
| Number of attainers on register (December 2012) | 1,473 |
| Canvass return rate (average 2010-2012) | 92.4% |
| Proportion of electors added in canvass (average 2010-2012) | 16.5% |

Results

Home movers

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|---------------------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DWP | 4,661 | 4,661 | 156 | 3% |
| Royal Mail | 4,348 | 4,348 | 101 | 2% |
| Both ¹¹⁰ | 68 | 68 | 5 | 7% |

¹⁰⁸ Relative to the 348 authorities in England and Wales (410 for 2001); highest to lowest proportions.

¹⁰⁹ This figure is taken from the 2001 census data as the 2011 figures are not yet available.

¹¹⁰ This category refers to individuals whose details were found on both databases used for that option e.g. found on both the DWP and Royal Mail databases for home movers.

Summary

Southwark sent registration forms out to all individuals identified on the data received, without matching the data against their electoral register or any other local dataset. This approach was in part a response to limited internal capacity, as a key member of staff had left shortly before the start of the pilot.

The proportion of new registrations achieved is among the lowest in the pilot. In part at least this will be due to Southwark's approach of writing to all records returned, as we know from other pilot areas' experiences that the data would have included incomplete addresses and out-of-date information.

The demographics in the area may have also played a part, with a relatively high level of population mobility contributing to the data received being even less current than in other areas.

Southwark received 1,209 responses to their write-out, however only 22% of these resulted in new registrations. The majority of the responses were from existing electors – 46% for the DWP data and 74% for the Royal Mail data. This is in line with other pilot areas, who found significant numbers of existing electors contained on the data returned to them.

Southwark found that it took a considerable amount of time to check the returned forms against their register and council tax records. They found the Royal Mail data harder to work with because it does not have UPRNs, thus making it more complicated to link the addresses on this data back to the addresses on their register.

Sunderland City Council

Demographic and electoral registration information

| Demographic information | | Rank ¹¹¹ |
|--|---------|---------------------|
| Population (16+) | 227,314 | - |
| Attainers (17 year olds) | 3,470 | - |
| Full-time students | 5.1% | 84 th |
| Private renters | 11.9% | 261 st |
| Estimated population mobility ¹¹² | 12.4% | 364 th |

Electoral registration data

| | |
|---|---------|
| Local government register entries (October 2012) | 216,445 |
| Number of attainers on register (October 2012) | 2,398 |
| Canvass return rate (average 2010-2012) | 89.8% |
| Proportion of electors added in canvass (average 2010-2012) | 8.4% |

Results

Attainers

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|---------------------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DfE | 353 | 341 | 18 | 5% |
| DWP | 4,197 | 3,920 | 227 | 5% |
| Both ¹¹³ | 1,913 | 1,878 | 239 | 12% |

¹¹¹ Relative to the 348 authorities in England and Wales (410 for 2001); highest to lowest proportions.

¹¹² This figure is taken from the 2001 census data as the 2011 figures are not yet available.

¹¹³ This category refers to individuals whose details were found on both databases used for that option e.g. found on both the DfE and DWP databases for attainers.

Home movers

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|---------------------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DWP | 12,815 | 12,356 | 946 | 7% |
| Royal Mail | 612 | 492 | 42 | 7% |
| Both ¹¹⁴ | 372 | 364 | 121 | 33% |

Summary

Sunderland found a relatively high number of duplicate records within the attainers file – around 1,000 of the DfE records seemed to be duplicates. They also noticed errors in addresses and apparent problems in the matching process for houses in multiple occupation and student residences.

On the home movers file, Sunderland found that 20% of the records related to existing electors, compared to 4% of the DWP records and 2% of the records found on both databases. On the attainers file, 5% of the total records related to existing electors. In line with other pilot areas' results, this suggests that the matching process for the Royal Mail data was less effective at excluding existing electors from the records returned.

Sunderland's results show a much higher registration rate for individuals whose details were found on both databases, indicating that these may tend to be the more accurate and current records.

¹¹⁴ This category refers to individuals whose details were found on both databases used for that option e.g. found on both the DWP and Royal Mail databases for home movers.

London Borough of Tower Hamlets

Demographic and electoral registration information

| Demographic information (2011 census data) | | Rank ¹¹⁵ |
|--|---------|---------------------|
| Population (16+) | 203,953 | - |
| Attainers (17 year olds) | 2,477 | - |
| Full-time students | 12.0% | 18 th |
| Private renters | 30.8% | 11 th |
| Estimated population mobility ¹¹⁶ | 22.5% | 18 th |

Electoral registration data

| | |
|---|---------|
| Local government register entries (December 2012) | 161,525 |
| Number of attainers on register (December 2012) | 1,531 |
| Canvass return rate (average 2010-2012) ¹¹⁷ | 85.8% |
| Proportion of electors added in canvass (average 2010-2012) | 21.3% |

Results

Attainers and students

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|--|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DfE and/or DWP and/or SLC ¹¹⁸ | 4,722 | 2,753 | 215 | 5% |

¹¹⁵ Relative to the 348 authorities in England and Wales (410 for 2001); highest to lowest proportions.

¹¹⁶ This figure is taken from the 2001 census data as the 2011 figures are not yet available.

¹¹⁷ Figures for the canvass return rate and proportion of electors added in canvass are only available for 2011 and 2012

¹¹⁸ Tower Hamlets did not report their results by database and for the new registrations also combined the results for the attainers and students file, meaning that we are unable to say which database the new registrations resulted from.

Home movers

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|----------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DWP | 9,317 | 8,183 | 632 | 7% |

Summary

Tower Hamlets received data for all three target groups of electors, over 55,000 records in total. Given this high volume, they decided to focus on records within the seven wards which had the lowest response rates to the 2012 annual canvass.

On the home movers file, Tower Hamlets started their follow-up with the most recently updated DWP records. In the timescales of the pilot, they were able to follow-up all DWP records updated within the past three months and some updated within the past six months. This means that they do not have any results for the older DWP records or Royal Mail records.

Tower Hamlets checked all records against their register and other council datasets before sending out registration forms. They found this to be a time-consuming process, and they had received a large number of records, which is why they did not complete the work on the home movers file within the pilot timescales.

Tower Hamlets found that 8% of the DWP home movers records related to existing electors, compared with 27% of the attainers and students records.

Wigan Metropolitan Borough Council

Demographic and electoral registration information

| Demographic information (2011 census data) | | Rank ¹¹⁹ |
|--|---------|---------------------|
| Population (16+) | 257,825 | - |
| Attainers (17 year olds) | 4,207 | - |
| Full-time students | 2.9% | 180 th |
| Private renters | 11.0% | 289 th |
| Estimated population mobility ¹²⁰ | 11.3% | 395 th |

Electoral registration data

| | |
|---|---------|
| Local government register entries (October 2012) | 245,390 |
| Number of attainers on register (October 2012) | 2,307 |
| Canvass return rate (average 2010-2012) | 92.9% |
| Proportion of electors added in canvass (average 2010-2012) | 9.0% |

Results

Home movers

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|--------------------------------------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DWP and/or Royal Mail ¹²¹ | 13,539 | 13,502 | 1,752 | 13% |

Summary

¹¹⁹ Relative to the 348 authorities in England and Wales (410 for 2001); highest to lowest proportions.

¹²⁰ This figure is taken from the 2001 census data as the 2011 figures are not yet available.

¹²¹ Wigan did not report their results separately by database.

The number of new registrations Wigan achieved was slightly higher than the pilot average for the home movers option.

Wigan checked the Royal Mail records against their electoral register before sending forms. They found that 16% of the records related to existing electors. They were unable to check the DWP records as the volume was too high for manual checks, which are time-consuming.

Wolverhampton City Council

Demographic and electoral registration information

| Demographic information (2011 census data) | | Rank ¹²² |
|--|---------|---------------------|
| Population (16+) | 200,047 | - |
| Attainers (17 year olds) | 3,471 | - |
| Full-time students | 5.5% | 80 th |
| Private renters | 13.2% | 208 th |
| Estimated population mobility ¹²³ | 12.8% | 347 th |

Electoral registration data

| | |
|---|---------|
| Local government register entries (October 2012) | 178,287 |
| Number of attainers on register (October 2012) | 1,713 |
| Canvass return rate (average 2010-2012) | 89.2% |
| Proportion of electors added in canvass (average 2010-2012) | 9.4% |

Results

Attainers

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|-------------------------------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DfE and/or DWP ¹²⁴ | 4,178 | 3,152 | 372 | 9% |

¹²² Relative to the 348 authorities in England and Wales (410 for 2001); highest to lowest proportions.

¹²³ This figure is taken from the 2001 census data as the 2011 figures are not yet available.

¹²⁴ Wolverhampton did not report their results separately by database.

Students

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|----------------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DfE and/or SLC | 2,945 | 2,206 | 430 | 15% |

Summary

Wolverhampton achieved a higher rate of new registrations on the students option than the attainers option. However, without knowing which database the new registrations came from it is difficult to reach any conclusions. For example, these 'student' registrations may in fact come from the DfE data which relates to school children aged 16-18 rather than students in higher or further education institutions.

Wolverhampton were able to conduct some of the checks using automated processes, but the checks against the register were conducted manually, which was a time-consuming process. During these checks, they found that 9% of the records on the attainers file and 8% on the students file related to existing electors.

Since submitting their results, Wolverhampton have completed an additional 85 new registrations, although they do not know which target group these relate to.

Wrexham County Borough Council

Demographic and electoral registration information

| Demographic information (2011 census data) | | Rank ¹²⁵ |
|--|---------|---------------------|
| Population (16+) | 109,026 | - |
| Attainers (17 year olds) | 1,626 | - |
| Full-time students | 4.3% | 100 th |
| Private renters | 11.9% | 258 th |
| Estimated population mobility ¹²⁶ | 12.4% | 361 st |

Electoral registration data

| | |
|---|---------|
| Local government register entries (December 2012) | 104,030 |
| Number of attainers on register (December 2012) | 868 |
| Canvass return rate (average 2010-2012) | 94.6% |
| Proportion of electors added in canvass (average 2010-2012) | 9.9% |

Results

Home movers

| Database | Sample size | Number of registration forms sent | Number of new registrations | % new registrations (from sample) |
|------------|-------------|-----------------------------------|-----------------------------|-----------------------------------|
| DWP | 2,191 | 2,035 | 190 | 9% |
| Royal Mail | 54 | 45 | 5 | 9% |
| Both | 20 | 16 | 0 | 0% |

¹²⁵ Relative to the 348 authorities in England and Wales (410 for 2001); highest to lowest proportions.

¹²⁶ This figure is taken from the 2001 census data as the 2011 figures are not yet available.

Summary

Wrexham discovered a lot of problems with the addresses on the data, including missing flat and house numbers and extensive abbreviations, making the data hard to use. Many of the addresses also used the previous county names, which have not existed since 1996.

Wrexham conducted a lot of checking and cleansing work before sending out registration forms, and found this to be a time-consuming task.

Wrexham's overall view was that data mining could be a useful tool but only if the quality of addressing was improved significantly.

The Electoral Commission
3 Bunhill Row
London EC1Y 8YZ

Tel 020 7271 0500

Fax 020 7271 0505

info@electoralcommission.org.uk

www.electoralcommission.org.uk

To contact our offices in Scotland, Wales,
Northern Ireland and the English regions,
please see our website.

We are an independent body set up by the
UK Parliament. We regulate party and
election finance and set standards for well-
run elections. We work to support a healthy
democracy, where elections and
referendums are based on our principles of
trust, participation, and no undue influence.

Putting voters first