Building the address register for the 2011 Census

A high quality, comprehensive list of addresses is fundamental to the 2011 Census. The address list provides the key to accurate delivery, collection and follow-up of questionnaires, as well as playing a central role in estimation.

Contents

The address register's context in the 2011 Census

In a change from previous censuses, in 2011 the questionnaires will be posted out to residential addresses, and as such the address register is right at the core of the 2011 Census process.

Communal establishments (for example: hotels, student halls of residence, caravan parks, nursing homes, prisons and hospitals) will have their questionnaires delivered and collected by hand, but postal delivery will dominate.

Questionnaires will be posted back by the public or, for the first time in this country, census answers will be collected from people completing their questionnaire online. In some areas that are known to have had a lower response in the past, questionnaires will be collected by field staff. In all cases, questionnaires will be tied back to the address list using barcodes unique to each questionnaire.

Addresses will obviously be used to enable posting out of questionnaires, but they will also be critical in tracking their return. Households from which a questionnaire has not been received will be visited by census collectors and encouraged to complete their census. The address register will be used to target and prioritise this follow-up process and it will contribute to estimation to take account of missing returns.

This central role of the address register is evident in 2011 Census quality targets. It is recognised that change and the complexity of the addresses means that no register can ever be 100 per cent complete. The aim, however, is for a register that is more than 99 per cent complete; to include at least 99 per cent of all existing residential addresses in England and Wales on Sunday 27 March 2011. At the same time, the register needs to have low levels of duplication, less than one per cent. This latter target is important, as duplication in the list will lead to wasted postage, potentially wasted staff hours through unnecessary follow-up and, critically, the risk of bothering householders who have returned a valid questionnaire.

Strategy for building and improving the residential address register

Match and validation of the national lists
No single national list currently provides a full solution to the 2011 Census requirement.

The core of the address register is formed by a match between the key national datasets - Royal Mail's Postcode Address File (PAF) and the National Land and Property Gazetteer (NLPG), maintained by Local Government. The version of PAF used in the match also incorporates grid references sourced from Ordnance Survey's MasterMap Address Layer 2 product. Both PAF and NLPG are high quality and improving products and are well suited to their intended uses. However, in ONS's view, neither provides a complete solution for the 2011 Census. Mismatches

between these two products (of around 5 per cent) suggest that ONS needs to draw addresses from both of these lists, as well as potentially from elsewhere.

Pulling together and de-duplicating these lists (around 27 million addresses each) is initially carried out using an automatic match between databases.

The ongoing matching process is being carried out on behalf of ONS by Manchester Geomatics Limited (MGL), while the infrastructure and management of the process is provided by Lockheed Martin UK. The NLPG / PAF match is supplemented by a match, using address identifiers derived from data provided by the Valuation Office Agency. This data does not provide extra addresses; rather, it helps to provide clues about links between the other products.

Together these matches provide a core of addresses – close to 95 per cent of the register - which are consistent between the two national products. These addresses are assumed to be valid.

The next stage focuses on those addresses which do not match, the mismatched addresses.

Intelligent Addressing, which maintains the NLPG on behalf of Local Government, provided further data cleaning as well as quality assurance for a match carried out by MGL in April 2009. This work cleaned out obvious problems before mismatched addresses were split up and provided for resolution to individual Local Authorities (LAs). Each local authority received a list of the remaining unresolved anomaly addresses for its area in November 2009. The numbers varied between a handful and several thousand for each local authority.

At the same time anomaly addresses were provided to Royal Mail who split up the remaining problem addresses and provided them to each of their 1,400 delivery offices. Whereas the work carried out by Intelligent Addressing was deskwork, Royal Mail carried out an almost exclusively field-based check, using postal delivery staff.

Following validation of different sets of anomaly addresses by the data suppliers and local authorities, ONS plans to carry out an extensive field check focused on areas where the register is believed to be weakest. This field check, employing around 400 staff nationally, will run between May and August 2010 and will cover around 15 per cent of addresses in the country, targeted using postcodes.

Postcodes containing a high level of mismatches between the source products have been targeted for the address check, as well as postcodes which are predicted to have high levels of multi occupation.

These two processes, matching between the key datasets and validating where there are differences, provide the core starting point for the register.

Keeping the address register up-to-date
The address register played a much less significant role in 2001 than it will in 2011, so its quality was much less critical. The list used in 2001 was clearly deficient in one serious respect however, as it was badly out of date. A cut was taken of the address register more than a year before it was used in the 2001 Census and so failed to pick up changes, and critically, new addresses that were introduced in the intervening period.

A key principle of the new strategy is that the central register will be kept up to date right up to census day. Updates will be taken for both the NLPG and PAF throughout this period and these will be applied to keep the central register up to date. This will provide a mechanism for picking up evidence of new addresses, most notably those included in the Local Land and Property Gazetteer (LLPG) updates provided by local authorities.

The main cut of the address register for post-out will be taken in November 2010, more than four months before census. Changes will be tracked during this period, as well as changes in the pipeline (such as buildings under construction) which will be identified and included where possible. More work is required to assess how this will work and what kind of evidence ONS will look for during this period, but it is intended that the register will only be days out of date rather than months.

Research, clerical work and using other sources
The first two stages outlined above provide the core of the address register and keep it up to date. It is clear that there is a great deal to be gained from further data cleaning within the register. On-screen inspection and comparison of records by a team of clerical staff will be used to validate any areas of the match where there doubt remains. This work will be supported using simple GIS software to plot addresses on maps and aerial photographs. Although potentially time-consuming, this type of work will be significantly more cost-effective than attempting to validate addresses in the field.

Work is also underway to investigate other sources of information that might help to refine ONS's decision about which addresses to post to and which to follow up. Third party sources on vacant and second homes, multi-occupation, house demolitions and information on utility meters, for example, are all of interest. In most cases these sources will not be used in their own right to add or remove addresses, rather to provide additional information to support decisions, or to help prioritise follow up. The section below on the evidence base provides further context for this.

Building an evidence base

On the surface it may seem that the address register will be built by merging the national datasets to form a single list and gradually knocking out spurious and duplicate addresses until a polished final list emerges. This is not a totally incorrect assumption, but the approach being taken is more flexible and much more powerful than this. Rather than simply discarding records, ONS will retain the full list of addresses throughout, and record the evidence provided at each stage against each address.

Figure 1 below provides a (simplified) example of the concept of storing evidence against each listed address.

This evidence base is built upon a list of all possible addresses, obtained by combining the key address sources, and shown here down the left hand side. Each row in the diagram relates to an individual address. Each column represents the views provided by each data supplier or obtained from an individual data source. In each case the code (eg **1L**) represents a view of whether this address is a valid residential address (**1**= residential address: send a questionnaire; **0**= not a residential address: do not send a questionnaire). The letter associated with each code provides an indicator of the evidence that supports this view. For example, a code of **0D** means 'Do not send a questionnaire: this is a duplicate record' while **1F** means 'Send a questionnaire: this address has been confirmed by fieldwork'.

Over time ONS will build up a stream of evidence from different sources, like an individual 'DNA', related to each address. It is then possible at any time, most notably when the final list is drawn, to use a series of business rules to decide which addresses will be sent a questionnaire.

Figure 1   The concept of an evidence base

Evidence	IA/LA NLPG	RM PAF	ONS FIELD	VOA CT ?	3rd ?	? ?
9 High Street	1L	1F	1R	1L	1L	1V
10 High Street	0C	1F	1R	0C	1L	1V
11 High Street	0S	1F	0X	1L	1L	0X
11a High Street	0X	1F	0X	0X	0X	0X
11b High Street	0X	1F	0X	0X	1D	0X
Flat 1/11 High Street	1L	0X	1R	1L	0X	1V
Flat 2/11 High Street	1L	0X	1R	1L	0X	1V

This approach provides ONS with virtually complete flexibility. It allows different sources of evidence to be balanced against each other and flexibility in deciding what level of certainty about an address is acceptable, depending on the circumstance. It also draws evidence from different sources at any time.

This approach also allows ONS more time to fully assess the relative quality of individual data sources and to make decisions on which forms of evidence are most reliable.

Strategy for building and improving the communal establishment address register

Communal establishments are defined for the census as 'managed residential accommodation'. They cover a wide range of types of accommodation, including prisons, nursing homes, hotels, university halls of residence, as well as more esoteric categories such as holiday camps and royal residences.

It is critical that the census properly captures information about such communal accommodation as some types can include large numbers of residents, or particular sub-populations that might otherwise be under-enumerated. In light of this a prioritisation of communal establishment types has been undertaken, based on importance to the census outputs, potential local impact on numbers and quality of current sources.

Preparing a list of communal establishments suitable for use in the 2011 Census is extremely challenging. There is a lack of coherent sources for many types of communal establishments and those sources that do exist (including PAF and NLPG) do not match with the census definition. Accordingly, early attempts to draw the communal address list directly from the matched national address list proved unsuccessful and the lists used in ONS's Pilot and 2009 Rehearsal have been below the standard required for the 2011 Census.

It is now clear that the communal establishment register requires a quite different approach than the one used for the residential list. Figure 2 provides an overview of current thinking.

Figure 2   Overview of Communal Establishment Strategy
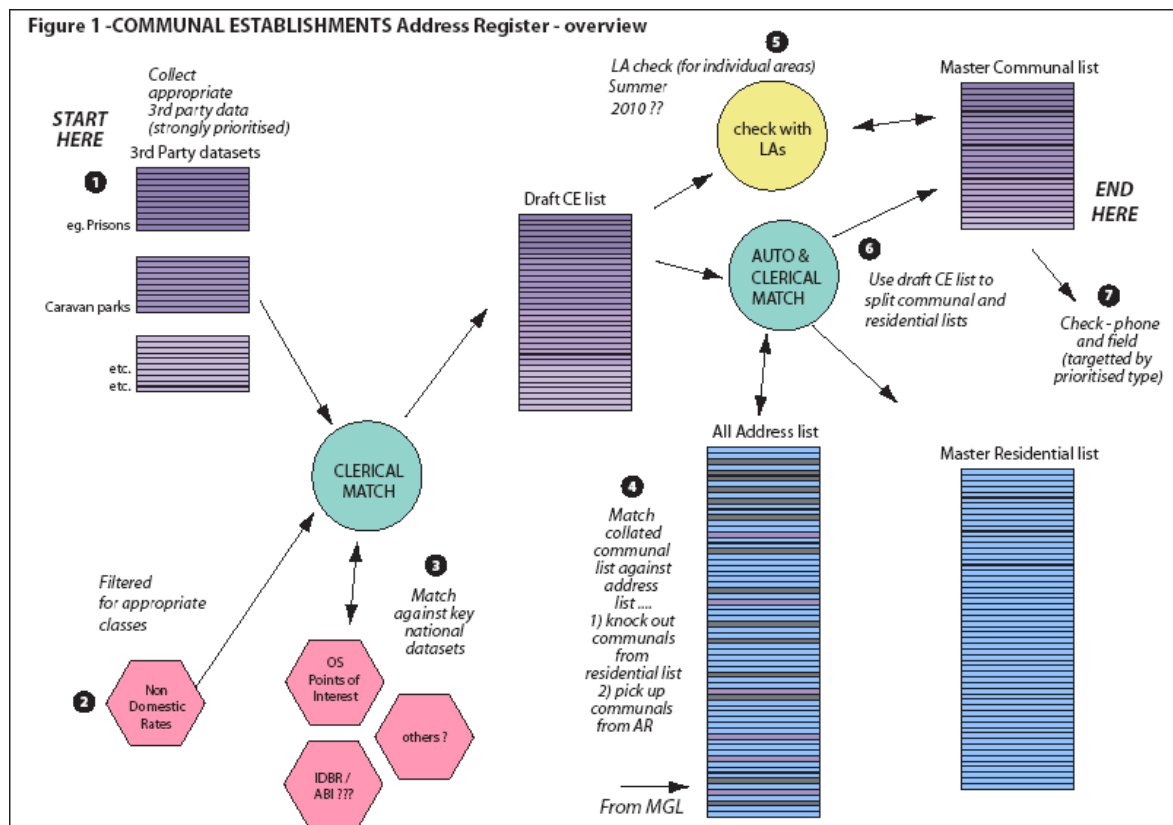
The starting point for the communal establishment list is the collection from third parties of lists for specific communal types. Many will be straightforward, but some much more difficult. Data found will be collated and matched against key national address lists (for example, from the Valuation Office Agency, Ordnance Survey's Points of Interest, etc).

The resulting full list of communal establishments will be matched against the final residential address register to ensure there is no double counting.

The potential significance of communal establishments for the census as a whole, and locally of some communal types (for example: caravan parks, halls of residence), means that a national check is appropriate. All communal establishments will be contacted (either by a field visit or telephone call) and details such as size and contact details collected.

Local authorities have a statutory responsibility for some communal types as well as local knowledge on the existence of particular establishments. Because of this local authorities will be asked to comment upon the list of communals for their area during the summer of 2010. Although local authorities will not be asked to sign off the list, their input is certain to be vital in helping to ensure the quality of the list.

Notes

Census Coverage Survey

The Census Coverage Survey (CCS), which takes place shortly after the census to check results and to support estimation, will not use the address register. The CCS is carried out as an exhaustive survey of households in defined areas and without an address list, and so will act as an independent validation of the quality of the register.

Households and addresses

A household, according to the census definition, is
- one person living alone; or
- a group of people living at the same address who share cooking facilities as well as a living room, sitting room or dining area.

Houses in multiple occupation (HMOs) are common and, increasingly, there are 'hidden' households, where several families with their own facilities live at a single address, without any external signs of multiple occupancy. The address check aims to identify such addresses and ONS is talking to pilot local authorities about how their own (HMO) data might feed into the census process.