Building the address register for the 2011 Census

Alistair Calder
Office for National Statistics

A high quality, comprehensive list of addresses is fundamental to planning for the 2011 Census. The address list will provide the key to accurate delivery, collection and follow-up of questionnaires, as well as playing a central role in fieldwork and estimation. This article provides a short summary of current thinking on developing the register, which will involve the address data suppliers, all local authorities, and many other stakeholders.[1]

Background

This article describes the approach being taken by the Office for National Statistics (ONS) to develop an address register to support the Census in England and Wales. The General Register Office for Scotland (GROS) and the Northern Ireland Statistics and Research Agency (NISRA) are responsible for the Census in Scotland and Northern Ireland respectively. Different circumstances and requirements lead to different approaches across the administrations, but regular contact and a common approach on key elements, such as outputs and disclosure control are helping to keep the approaches aligned.

The 2011 Census project in England and Wales consists of four main stages. These are:

- The Census Test, carried out in September 2007 in five local authorities to test major innovations
- The Census Rehearsal, run in October 2009 in three different local authorities – Lancaster, Anglesey and Newham – to test near-final field and operational procedures and systems
- The 2011 Census itself – Census Day is Sunday 27 March 2011
- The Census Coverage Survey, in May 2011, to assess the coverage of the 2011 Census and to allow accurate estimation and adjustment for under-coverage

There have also been a number of small-scale tests conducted in selected areas, to inform how we will deal with particular community or population groups, or to carry out certain pre-Census processes, such as the checking of addresses.

The importance of addresses in the 2011 Census

The address register is right at the core of the 2011 Census process.

In a change from previous censuses, in 2011 the vast majority of questionnaires will be posted out to residential addresses. Around five per cent of forms will be hand delivered in those areas that are known to be most difficult to enumerate, and to communal establishments (for example: hotels, student halls of residence, caravan parks, nursing homes, prisons and hospitals), but postal delivery will dominate.

Questionnaires will mainly be collected by the public posting them back or, for the first time in this country, by completing the questionnaire via the internet. In some areas that are known to have had a lower response in the past, questionnaires will be collected by enumeration staff, or information will be taken over the telephone. In all cases, forms will be tied back to the address list using barcodes unique to each form.

Addresses will obviously be used to enable posting out of questionnaires, but they will also be critical in tracking their return. Those households, from which a questionnaire has not been received, will be visited and encouraged to complete the form. 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 our quality targets. It is recognised that change and the complexity of the addresses mean that no register can ever be 100 per cent complete. We are, however, aiming for a register that is more than 99 per cent complete, that is, that it includes at least 99 per cent of all existing residential addresses in England and Wales on Census Day. 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 contacting householders who have already returned a valid form.

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.

The section below outlines the three main streams of the strategy currently being developed to deliver the address register.

Strategy for building and improving the address register

Stream 1: Match and validation of the national lists

No single national list currently provides a full solution to the 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 of high quality and improving products and are well suited to their intended uses. However, in the view of ONS, neither of these products provides a complete solution for the 2011 Census. Mismatches between these two products (of around five per cent) suggest that we do need 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 matching process is being carried out on ONS's behalf 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 with 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 stages focus on those addresses which do not match, the anomaly addresses.

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

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

Following validation of different sets of anomaly addresses by the data suppliers and LAs, ONS plans to carry out an extensive field check focussed upon areas where the register is believed to be weakest. This field check, employing around 400 staff nationally, will cover around 15 per cent of addresses in the country, targeted using postcodes. The precise criteria for identifying postcodes with high levels of potential problem addresses are still being investigated, but a mismatch between the source products and the predicted level of multi-occupation are likely to be the main factors.

Box one

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 are common and, increasingly, there are 'hidden' households, which would involve several families with their own facilities living within a single address, without any external indication of multiple occupancy. The address check, discussed later, will aim to identify such addresses.

Other data sources are being used to find areas that are more likely to have more multi-occupation, in order to target hand delivery.

Households, then, do not always equate to addresses – but a high quality list of residential addresses and, as far as possible, households, will be at the very centre of the census process in 2011.

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

Stream 2: 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, 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 us with a mechanism for picking up evidence of new addresses, most notably those included in the Local Land and Property Gazetteer (LLPG) updates provided by LAs.

The main cut of the address register for post-out is taken on 11 November 2010, more than four months before Census. We will, however, continue to keep track of changes during this period and also specifically look for direct intercept of change in the pipeline. More work is required to assess how this will work and what kind of evidence we will look for during this period, but it is intended that the register will only be days out of date rather than months.

Stream 3: Research, clerical work and using other sources

The first two stages outlined above provide the core of the directory and keep it up to date. It is clear that there is a great deal to be gained from further data cleaning within the directory. 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 remains doubt. This work will be supported by 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 our decision about which addresses to post to, and to follow up. Third party sources on vacant and second homes, multi-occupation, house demolitions, and information on utility meters, for example, will all be of interest. In most cases these would not be used in their own right to add or remove addresses, rather than to provide additional information to support decisions, or on prioritising 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 removing spurious and duplicate addresses until a final list emerges. This is not totally incorrect, but the approach being taken is more flexible and much more powerful than this. Rather than simply discarding records, we will retain the full list of addresses throughout, and record the evidence provided at each stage against each address.

Figure 1 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

Figure 1	The concept of an evidence base					
Evidence	**IA**	**RM**	**F**	**LA**	**3rd**	**?**
9 High Street	1L	1F	1R	1L	1L	1G
10 High Street	0C	1F	1R	0C	1L	1G
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	1G
Flat 2/11 High Street	1L	0X	1R	1L	0X	1G

column represents the views provided by each data supplier or obtained from an individual data source. In each case the code (for example, **1L**) represents a view of whether this address is a valid residential address (**1** = residential address – send a form, **0** = not a residential address – do not send a form). 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 form – this is a duplicate record' while **1F** means 'Send a form 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 we have to draw the final list, to use a series of business rules, to decide which addresses merit being checked or sent a form.

This approach provides us with virtually complete flexibility. It allows us to balance different sources of evidence against each other and to calibrate the degree of certainty we require for specific applications. It also allows us to draw evidence from different sources at any time.

Critically, this approach also allows us further time to fully assess the relative quality of individual data sources, and to make decisions on which forms of evidence are most reliable.

As we move nearer to the 2011 Census, we will want to develop and test a core set of addresses, with help from the system of building and holding evidence against individual addresses, rather than just removing suspect addresses.
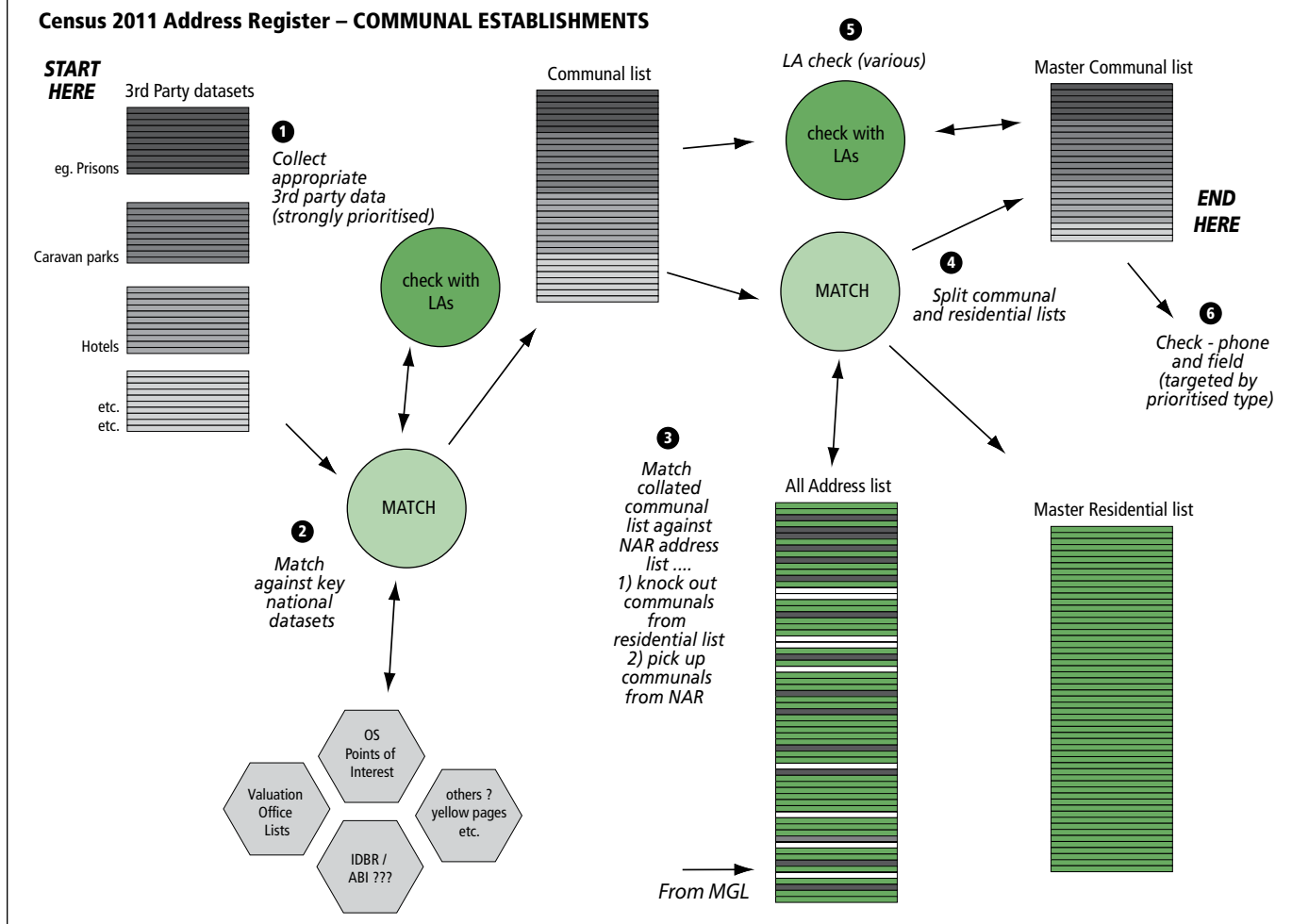
Communal establishments

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 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.

Preparing a list of communal establishments suitable for use in Census is extremely challenging. There is a lack of coherent sources for many types

Figure 2	Overview of communal establishment strategy

Census 2011 Address Register – COMMUNAL ESTABLISHMENTS

of communal establishments, and those sources that do exist (including PAF and NLPG) do not match with the Census definitions. Accordingly, early attempts to draw the communal address list directly from the matched national address list proved unsuccessful, and the lists used in our Pilot and Census Rehearsal have been below the standard we aim to obtain for the 2011 Census.

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

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 we are able to find will be collated and matched against key national address lists (for example, from Valuation Office Agency, Points of Interest, etc).

The resulting full list of communal establishments will be matched against the final National Address Register in order to ensure that there is no double counting – the final result being separate residential and communal lists.

The communal establishment list will be checked in the field at the same time as the residential address list. For those parts of the country covered by the address checking process all communal addresses will be checked and details and size and contact details collected. The potential significance 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. Work is currently underway to develop a prioritisation of communal establishment types, based on importance to the Census, potential local impact on numbers, and quality of current sources.

It is increasingly obvious that local authorities have a vital role in helping us develop a quality list of communal establishments. LAs have statutory responsibility for some communal types. In many cases local knowledge will be invaluable. As far as possible LAs will be provided with an opportunity to validate the lists of communal addresses in their areas.

Next Steps

It is felt that the strategy as outlined above offers ONS the best solution for both residential and communal addresses.

Work continues on developing the strategies and on refining the approach to address checking, to be carried out between May and August 2010.

All of the plans as outlined, however, are subject to change in the light of the ongoing programme of research, and our findings from the Census Rehearsal, but also in the light of discussions with stakeholders. The views of local authorities in particular were taken at a series of eleven regional Address Register Workshops, held in September and October 2009, and through ongoing liaison as plans develop.

Key findings

- A high quality, comprehensive list of addresses will play a central role in the 2011 Census. The address list will provide the key to accurate delivery, collection and follow-up of questionnaires, as well as playing a key role in fieldwork and estimation.
- The lack of a single national register suitable for Census purposes means that ONS are currently pulling together a new register drawing upon a range of sources – most notably the Postcode Address File (PAF) and the National Land and Property Gazetteer (NLPG).
- 'Anomaly' addresses that do not match between these products are being verified with data suppliers, with local authorities and via an ONS field check of around 15 per cent of the country.
- The resulting register will be kept up to date right up until Census Day – 27 March 2011 – through regular updates from the source addressing products. Over time a stream of evidence is being built up about each possible address and this will be used to make decisions about which addresses to send forms to.
- A different approach is being taken for addresses for 'communal establishments' as these have quite different characteristics from residential addresses. Communal Establishments, however, are particularly important for the Census as they may contain significant 'hard-to-count' populations so this is an area of particular focus at present.
- Development of the register is underpinned by a programme of research and regular liaison with local authorities and other key stakeholders. Comments on the approach being taken or suggestions for improvements are welcomed at any time.

Views or comments on any aspect of the strategy as outlined in this article, and particularly suggestions on improvements to the adopted approach, are strongly welcomed at any time.

Please contact alistair.calder@ons.gov.uk with any views.

Notes

1 It is important to realise that the register currently being developed is solely for use in the 2011 Census. The address register being developed could have a wider role as a rolling address or population register, but this is not the focus of the current work. There are currently no plans to make the register publicly available and there are significant commercial, licensing and confidentiality obstacles that would have to be cleared before this would be possible.