

**Economic and Social Council**

Distr.: General  
13 March 2012

English only

---

**Economic Commission for Europe****Conference of European Statisticians****Group of Experts on Population and Housing Censuses****Fourteenth Meeting**

Geneva, 24-25 May 2012

Item 4 of the provisional agenda

**Address/dwelling listing****Building the address register for the 2011 Census – England and Wales****Note by the Office for National Statistics, United Kingdom***Summary*

A high quality, comprehensive list of addresses was fundamental to planning for the 2011 Census in England and Wales. The address list provided the key to accurate delivery, collection and follow-up of questionnaires, as well as playing a central role in fieldwork and estimation. This paper describes the approach taken and summarises the main challenges and lessons learnt.

Note: This paper describes the situation in England and Wales. The General Register Office for Scotland (GROS) and the Northern Ireland Statistics and Research Agency (NISRA) were 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 helped keep the approaches and key outputs aligned.

**I. The role of the address list in the 2011 Census**

1. This document was prepared by Alistair Calder, Office for National Statistics (ONS).
2. The address register was right at the core of the 2011 Census process in England and Wales. In a change from previous censuses, in 2011 questionnaires were posted out to

every residential address. Questionnaires were hand delivered and collected only for communal establishments ('managed residential accommodation'). Questionnaires were mainly returned by the public posting them back or, for the first time in the United Kingdom, by completing the questionnaire via the internet.

3. Addresses were obviously used to enable posting out of questionnaires, but they were also critical in tracking their return. Each questionnaire was bar-coded and scanned on return. Every household, from which a questionnaire had not been received, was visited by an enumerator and encouraged to complete the form. The address register was used to target and prioritise this follow-up process and it also contributed to estimation to take account of missing returns.

## **II. Key requirements**

4. The key requirements for the address register were:

- Completeness : ONS required a complete list of house spaces (essentially anywhere where someone might be living) for England and Wales;
- Minimal under and over coverage;
- Coverage of, and distinction between, both residential household addresses and 'communal establishments'

5. A household, according to the Census in England and Wales, 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.

6. 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, was aimed at identifying such addresses.

## **III. Quality targets**

7. The central role of the address register was evident in our quality targets. ONS required a register that had less than 1 per cent under-coverage, that is it should include at least 99 per cent of all existing residential addresses in England and Wales on Census Day (27 March 2011).

8. At the same time, the register needed to have low levels of over-coverage (errors, wrongly coded commercial premises, demolitions etc); the target was less than 2.5 per cent over-coverage.

9. In particular ONS was aiming for less than 1 per cent duplication of addresses (within the over-coverage) - where there were multiple versions of the same address, perhaps differently described. This latter target was important, as duplication in the list would lead to wasted postage, potentially wasted staff hours through unnecessary follow-up and, critically, the risk of bothering householders who have returned a valid form.

10. The Census Coverage Survey (CCS), which aimed to check results and to support estimation, took place around two months after the Census but did not use the address register. The CCS was carried out as an exhaustive survey of households in defined areas

and without an address list, and so acted as an independent validation of the quality of the register.

11. Given its key place underpinning the Census, development of the address register was one of the main areas of research leading up to the 2011 Census. Although there were a number of high quality, and to some extent competing, national address lists, early research showed that no single product came close to meeting the specific needs of the Census. ONS were forced to construct a new composite address register by pulling together the best parts of the national sources. This section provides a summary of our strategy for carrying out this work.

## **IV. Building the residential list**

12. The approach to developing the residential household address register had five stages.

### **A. Matching**

13. The core of the Address Register was 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 match also incorporated grid references from Address Layer 2 (AL2) maintained by the Ordnance Survey (the national mapping agency).

14. Automated address matching was carried out on behalf of ONS by Lockheed Martin who sub-contracted the work to Manchester Geomatics Limited (MGL). The process was as follows:

- The address lists were standardised to a consistent address structure. Address classifications were also standardised in order to facilitate differentiation between classifications.
- AL2 and PAF were merged using unique address identifiers. These two products were then treated as a single address list in all subsequent steps. The intention here was to maximise the benefit of precisely geo-coded and classified addresses from AL2 with the more up to date information on PAF from Royal Mail.
- The addresses from NLPG and PAF/AL2 were then textually matched using a series of algorithms including exact and fuzzy matches.
- The remaining addresses that could not be textually matched were then spatially matched.

15. This process matched about 94.5 per cent of addresses nationally, although the match-rate did vary between local authorities – with a handful of authorities with match rates between 70 and 80 per cent. Generally, the match rates were lower in London than other areas (averaging 90 per cent) probably due to higher numbers of flats which did not fully match on the source products.

16. These core matched addresses which were consistent between the two national products (and classified as residential) were initially assumed to be valid (a decision based on evidence from the census rehearsal).

17. Throughout the development of the address register, this matching process was carried out 16 times – refining the match and validating the approach each time. Both our

processes and the source products themselves were constantly improving up until the time of the Census.

18. Together these matches provided a core of addresses –close to 95 per cent of the register- which were consistent between the two national products. These addresses were assumed to be valid. The next stages focus on those addresses which do not match, the anomaly addresses.

## **B. Anomaly resolution by data suppliers**

19. The remaining, unmatched, addresses were reviewed to resolve anomalies and over 500,000 were then sent to local authorities or Royal Mail for checking and resolution. Over 90 per cent of local authorities participated in this exercise as did Royal Mail staff across England and Wales. This work ultimately led to around 40 per cent of these anomaly addresses being identified as invalid and excluded from our list.

## **C. Field check**

20. In the summer of 2010, ONS conducted a field check of 3.6 million addresses. The Address Check ran from May to August 2011 covering 15 per cent of the addresses in England and Wales. ONS targeted those postcodes (average size around 15 addresses) that contained the highest numbers of unmatched addresses or two or more multi-occupancy addresses (a particular concern for the Census). The check employed 240 people and primarily checked residential addresses.

21. Address Checkers were provided with pre-printed lists for whole postcodes (not just the anomaly addresses). The work was split into 28,000 separate workloads - each with an average of around 10 whole postcodes. They were required to check these lists as well as looking for new household spaces not included. Special address checkers (a further 85 staff distributed nationally according to the distribution of communal establishments) visited thousands of the more complex communal establishments including all universities, prisons, hospitals and army camps.

22. This exercise identified some missing addresses and identified nearly half a million addresses to remove from our list. Quality control measures were put in place to assess the quality of the address checkers' work. Addresses were removed from the address lists provided to address checkers (but not the final address register), then the quantity subsequently found as 'new addresses' was measured. The addresses removed were a 1 per cent random sample of matched addresses. Address checkers found 77 per cent of these removed addresses, which showed that although field checking finds missing addresses, it does not deliver totally correct lists.

23. The address check was completed to time and under budget and is considered to have been highly successful. As well as validating the address list in areas with large numbers of anomalous or complex addresses, the check provided a real insight into the quality of the matched data and the other sources of evidence available.

## **D. Building Evidence**

24. The information from the resolution work and ONS's own checking was collated into an evidence base - pulling together everything known about each address. The use of multiple sources for creating the address register led to difficult decisions on what to include and exclude. Business rules were applied but ONS took a cautious approach;

wherever there was doubt, addresses were included rather than excluded - minimising under-coverage was the over-riding objective. (See further discussion below.)

## **E. Building the final list**

25. All of the above work was integrated to form the final list. The majority of addresses (25.1 million) were supplied for printing onto questionnaires in October 2010. A second list of nearly 300,000 addresses was compiled in January 2011 incorporating subsequent changes and also addresses likely to be built and occupied between then and Census Day. One of the main criticisms of the 2001 Census was its use of an address list that was three years out of date; this time ONS ensured the list was as current as possible.

## **V. Balancing the evidence**

26. The principle of building the address register was conceptually simple. Those addresses which matched between the two datasets were considered to be valid. Those that did not match were considered valid unless there was evidence to the contrary (although this has potential to result in unnecessary duplication). By gradually refining and improving the match the number of anomalies and the level of duplication were reduced. By building evidence about and knocking out erroneous anomaly addresses ONS improved quality and reduced over-coverage.

27. Since it was critical that ONS avoided under-coverage (that is, missing valid addresses) those addresses which existed only in one or other of the datasets were included in the list to be sent a questionnaire – except where were ONS had clear evidence that they were in error. Accordingly a census form was sent to all addresses included in the NLPG and all addresses included in the PAF – except where there was strong evidence to suggest that an address did not exist – or was a duplicate of another.

28. Although the principle was fairly simple there remained one key issue - the anomaly resolutions and evidence available to ONS were often contradictory. Inevitably, the Royal Mail normally believes its own version of an address to be correct – while LAs favour the NLPG - while both versions of the address may be equally valid.

29. Accordingly ONS could not base a decision to include or exclude a particular address upon a single piece of evidence. Rather it retained the full list of addresses throughout and recorded the evidence provided at each stage against each address.

30. Figure 1 provides a (simplified) example of the concept of storing evidence against each possible address. Each code represents the view of an individual party on a given address.

31. Over time ONS built up a stream of evidence – an individual ‘DNA’ related to each address. It was then possible at any time, but most notably when ONS prepared the final list – to use a series of business rules to decide which anomaly addresses merit being checked or sent a form.

**Figure 1**  
**The concept of an evidence base**

Evidence	IA/LA IILPG	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

32. Obviously implementation of this idea was more difficult than this principle implies – and the rules used were exceptionally complex in some places. However the concept of building evidence for later application to the matched set provided real flexibility - allowing ONS to balance different sources of evidence against each other and to calibrate the degree of certainty we required for specific applications. Critically this approach also allowed time to fully assess the relative quality of individual data sources and to make decisions on which forms of evidence were most reliable.

33. Business rules were built up and refined over time – first from logic and then based upon evidence gained in the field. This approach was considered a significant success.

## VI. Keeping the register up to date

34. The address register played a much less significant role in 2001 than in 2011 so its quality was much less critical. The list used in 2001 was clearly deficient in one serious respect however – it was badly out of date. A cut was taken of the address register as early as three years before it was used in Census(!) and so failed to pick up changes, and critically new addresses, introduced in the intervening period (although the method of using hand delivery and field checking of the list in 2001 made the problem much less significant).

35. One of the key issues here is the high degree of change taking place in all of the source products. New addresses are being constantly added to the separate lists – at a rate of around 30/40 thousand new addresses a month. Add to this a similar number of deletions and of changes to addresses and the issues in keeping the lists up to date and in sync are obvious.

36. A key principle of the new strategy was that the central register was kept up-to-date right up until Census Day. Updates were taken for both the NLPG and PAF throughout this period and these were applied to keep the central register up to date. Critically this provided ONS with a mechanism for picking up evidence of new addresses – most notably those included in the Local Land and Property Gazetteer (LLPG) updates provided monthly by LAs.

37. In order to start the printing of the envelopes the main cut of the address register had to be taken more than six months before Census. However, ONS continued to keep track of changes during this period and produced a supplementary list for printing in February 2011. Likely future changes were included by adding change in the pipeline (candidate addresses from builders, not yet built etc). A decision was made to include in a large number of addresses where there was only weak evidence of existence – better to risk some over-coverage than to miss any new properties.

The final register was only *days* out of date rather than *years*.

## VII. Building the communal establishment address register

38. A separate list of communal establishments (places providing managed residential accommodation such as care homes, army camps and prisons) was required as these were enumerated using different questionnaires and methods from those for households.

39. Different sources were used to develop a list of communal establishments. Instead of national address products, some 50 specialist source lists for different types of establishment were used. However, the same approach to determine the contents of the final list was adopted - assessing evidence for each address including that supplied by local authorities to determine which ones were included. Most of the matching was carried out manually and clerically because of the complexity of these addresses. Again, ONS took a cautious approach and included addresses where there was doubt about their existence and/or whether they had any residents.

40. Producing the list of communal establishments was one of the most challenging parts of the address register project. Duplication between the communal and residential lists proved to be a problem in places. Although the end result was acceptable, in retrospect ONS should have started earlier and kept the residential and communal lists better aligned.

## VIII. Assessment of quality

41. Measuring against the targets outlined before, the residential household list:

- is thought to have met the target of under-coverage not exceeding 1 per cent; ONS estimate that it achieved 0.8 per cent. Furthermore, there was no evidence of small area problems with missing addresses;
- did not meet the target for over-coverage not exceeding 2.5 per cent; ONS estimate it was 3.3 per cent. This was considered acceptable and to some extent the result of conscious decisions – the under-coverage target was always paramount; and
- far exceeded the target for 99.0 per cent for positional accuracy; ONS estimate it was 99.7 per cent.
- In terms of communal establishments ONS identified a small number of new sites during the operation (1.9%) that had been missed from the original list. Over-coverage was 27 per cent, due to a combination of:
  - caution in construction of the list: for example ONS included sites where there was a chance of usual residents (hotels, caravan parks etc) knowing that many would have none; and
  - duplication with the household list

42. The higher than hoped for levels of duplication both within the household list and between households and communal establishments did cause issues for other parts of the

Census operation (the field operation, data processing and contact centre) and inconvenience to some members of the public. This was regrettable but was ultimately a price worth paying for minimising under-coverage.

## **IX. Key conclusions and lessons learned**

- In principle the approach of matching address products together, checking anomalies and applying evidence would seem to have been the correct one – although there are obvious areas for improvement the plan worked.
- Developing an evidence base and applying rules led to high coverage but caution in the rules did lead to higher than hoped for over-coverage and duplication. In retrospect ONS should have applied more resource to perfecting the business rules.
- Although the over-coverage target was not met, this was a conscious decision at the time (to include addresses if in doubt). The advantage of meeting the under-coverage target far outweighed the minor added cost of over-coverage in the field.
- ONS could not have delivered such a high quality address register without the support and commitment of address suppliers and local authorities. Partnership with the data suppliers was absolutely critical and a key success story from the project.
- The household and communal establishment lists should be developed together to minimise duplication

43. Although ONS can spot potential improvements for the future, in retrospect the address register exceeded its most important target of achieving high coverage – it was estimated that included 99.2 per cent of existing households. On balance this challenging project which provided the bedrock of the Census is felt to have been a considerable success.

---