



CENSUS ADVISORY GROUP

AG (07) 08

Household Frame development

Introduction

1. This paper reports further on the approach (earlier described in Paper AG (06)03 that ONS are taking to develop a Household Frame for the 2011 Census in England and Wales. It outlines the two key areas of development required to ensure a Household frame is available, of sufficient quality, to support the current Census design:
 - creating a base address list; and
 - conducting an address checking exercise 6 months before the Census.
2. Advisory Groups are asked to note the approach to developing a Household Frame for the 2011 Census.

Base address list

Timing and approach

3. A final decision on the 2011 approach to create the initial address list will be made in March 2009, on the basis of phase 2 research of the address products, and ongoing discussions with suppliers about licensing costs. The proposed strategy for 2011 is to provide an address list sourced from both AL2 (Ordnance Survey's MasterMap Address Layer 2) and NLPG (National Land and Property Gazetteer), with a methodology to be researched during phase 2.

Background

- 4 Without a single definitive national address register for England and Wales, Census needs to assess the two competing address products, Ordnance Survey's MasterMap[®] Address Layer 2 (AL2) and the National Land and Property Gazetteer (NLPG) for the 2011 Census. Other supplementary address data sources are needed for communal establishment and special enumeration addresses. The products will be assessed against a range of criteria in the Census Household Frame requirements.
- 5 AL2 is sourced from OS's ADDRESS-POINT[®], which in turn is derived from the Royal Mail[®] Postcode Address File (PAF[®]). It has been further enhanced by new datasets –

Objects without postal addresses (OWPAs) and Multi-occupation without postal addresses (MOWPAs), and is now integrating Royal Mail's Multi-Residence file.

- 6 NLPG is collated from Local Land and Property Gazetteers (LLPGs), which have been created by all councils with a statutory street naming and numbering function. LLPGs were created by consolidating various local authority address data sources, such as Council Tax and non-domestic rates, Electoral Roll and Planning data.
- 7 Phase 1 of the research focused on the coverage of AL2 and NLPG in the Census Test areas during the address checking exercise run in September/October 2006. The research showed that neither product fully meets Census requirements. The full Report of the Phase 1 research - *Coverage of address registers for 2007 Census Test - Phase 1* is available on the ONS website at: <http://www.statistics.gov.uk/census/2011census/CollectingtheInfo/inputgeography.asp>
- 8 The Executive Summary from the Report is attached at Annex A.

Phase 1 research – outcome

- 9 The research showed that both registers were incomplete at the time of the address checking exercise, with significant under-coverage in Camden and Liverpool, particularly for multi-occupied addresses (flats). NLPG has better coverage in Camden and Liverpool AL2 (see table below); but AL2 has better coverage in the other three less built-up areas, in Stoke, Bath and NE Somerset, and, significantly so, in Carmarthenshire

Test area	AL2 addresses as % of those found in field	NLPG addresses as % on those found in field
Bath and NE Somerset	94.1	93.1
Camden	89.4	96.0
Carmarthenshire	96.4	84.1
Liverpool	94.4	100.6
Stoke-on-Trent	98.6	97.9

- 10 Both products failed to classify a number of residential addresses, and wrongly classified addresses that were found to be non-residential by the address checkers. As supplied, NLPG failed to classify 20,000 addresses at all. Although these records were excluded from the residential address counts, they were used to match against those addresses the address checkers confirmed existed.
- 11 It should be noted that, as the areas known to be harder to enumerate were over represented in the Census Test areas, the coverage figures cannot be applied nationally.
- 12 The methodology and findings were externally assured by Manchester Geomatics.

On-going development programmes

Both address suppliers have programmes for developing/improving their products over the assessment period, and beyond. A key enhancement over the next few months is the adoption by both suppliers of the “BS7666: 2006 Spatial datasets for geographical referencing” standard. This will ensure the data supplied is better presented and formatted, and also ensures that quality reports and metadata are provided about the datasets.

Phase 2 research

13 Phase 2 of the research (**August 2007-March 2009**) has started with a broader scope and geographical coverage. The prime aims of the phase 2 research are to:

- achieve the ‘best’ list possible to feed into address checking for Dress Rehearsal, and 2011; and
- ensure key stakeholders (including LAs) are sufficiently content with the addressing strategy.

14 More broadly, the scope of the research includes evaluating the following topics, split into key and secondary:

Key topics

- **national coverage** – adopting a suitable methodology to support national coverage;
- **positional accuracy** – high resolution grid references with an indicator of how it has been assigned;
- **multiple occupancy** – evaluating multiple occupancy in the address registers;
- **Welsh language** – addresses are available in Welsh language;
- **licensing/costs** – establishing the terms and conditions of acquiring the address products;

Secondary topics

- **data supply and currency** – data is available and supplied on customer specified dates;
- **‘soft’ issues** – overall ONS/supplier relationship and their ability to respond to queries;
- **data loading tools** – assess software products that will enable data to be loaded into the ONSG Oracle and SQL Server database;
- **time series** – analysis on improvement over time focusing on coverage, classifications, positional accuracy and quality;
- **linkage to buildings** – all addresses linked to a building, either physically or via unique reference;
- **data linkage** – ability to link address products to other data sources;
- **BS7666** – compliance of the address products to BS7666 (2006) standards;

- **occupancy type/status** – ability to identify whether an address is private or communal (type) and vacant or occupied (status);
 - **historic data** – ability to identify historic/demolished addresses;
 - **data and product improvements** - evaluate how quickly on the ground changes are reflected in the product, and how effectively products have improved as planned;
 - **capacity estimates** (relating to Communal Establishments only) – evaluate capacity/bed space estimates received from data suppliers.
- 15 The bulk of the Phase 2 research will be to develop a methodology for comparing the two national lists and identifying anomalies, and then carrying out field checks in selected areas to determine ‘the reality’ on the ground. This will result in:
- (a) an understanding of the quality of the two products;
 - (b) a decision on whether to use a single product or a combination of the two products; and
 - (c) a potential methodology for combining the two products, if required.
- 16 Key to the research will be gaining buy-in to the approach and the findings from key stakeholders (including OS, IDeA, and LAs) so that the final choice of address product has the best possible chance of being supported whilst still meeting our quality requirements.
- 17 Detailed methodology is currently being developed and will be quality assured by a sub-group of the UK Census Design and Methodology Advisory Committee (UKCDMAC).

Address checking in 2011

- 18 As well as the work to develop a good base address list, it is clear that a full address check in England and Wales is required for the 2011 Census to fulfil Census requirements for a Household frame. The evaluation of the two products using the information from the Test showed that neither was of a sufficient quality to confidently identify areas that would not require an address check. Therefore it is planned to do a 100 per cent address check prior to the 2011 Census.
- 19 As a result of the evaluation of the 2007 Test and other considerations it has been decided that the approach to address checking in 2011, will broadly be the same as in the Test. Address checkers will have one list for England and Wales, the base address list supplied by ONS Geography, and will check every address on the ground whilst at the same time search for, and identify, addresses missing from the list.
- 20 As in the Test, there will be two types of address checking:
- **discretionary contact** – in a discretionary contact area address checkers will be advised to make contact with the household at their discretion, for instance when they see more than one doorbell on a door listed as a single address; and

- **full contact** – in a full contact area address checkers will be required to make at least three attempts to make contact with every address.

21 ONS will be researching how to target the full contact method in a more intelligent way, specifically looking at:

- comparing the two national products, to identify areas (postcodes, Output Areas (OAs), Lower Super Output Areas (LSOAs) where the two lists differ as an indicator for that area requiring full contact;
- identifying areas of high change in the address lists, potentially through other data sources and/or through LA identification; and
- identifying areas of high-multi-occupation, which tends to be where there is the most success with the full contact method.

22 Additional improvements are also being researched for address checking:

- **Moving to a six-month long address check.** For the Test this was done over a *six-week* period, and the address checkers found some 12,000 addresses resulting in a mammoth task to key the addresses and match in order to update the address register for questionnaire printing. Updating the national address register in this time scale and this method in 2011 is untenable. Therefore it is intended to move to an address check lasting some 4-6 months, with completion still approximately six months before Census day. Moving to an address check over a longer period in this way will provide more time to update the address register accurately and should improve the quality of address checkers by providing more scope for improving their skills.
- **Training.** Based on the number of new addresses identified during enumeration that were believed to exist at the time of the address check but which were missed, the evaluation of the Test indicated that there is room for improvement in the training and procedures. As a supplement, the use of controlled errors (addresses manually removed by ONS to enable field managers to validate the address checkers work) during address checking suggested that a significant proportion of controlled errors were not found.

Coverage of address registers for 2007 Census Test - Phase 1 -
EXECUTIVE SUMMARY

Fundamental to a successful Census is a complete and accurate address frame to support enumeration.

Two address registers, OS MasterMap® Address Layer 2 (AL2) and the National Land and Property Gazetteer (NLPG) are being evaluated by the Office for National Statistics (ONS) to determine their suitability to support the 2011 Census and 2009 Census Rehearsal. ONS is looking for product(s) / methods which are likely to give best overall coverage for the Census.

Phase 1 of the research was an important but interim evaluation of the position in October 2006. An address-checking field exercise was conducted in September 2006, making it possible to compare these results with the two address registers.

The purposes of phase 1 were, firstly, to assess the position as in October 2006 and, secondly, to provide recommendations for overcoming the problems found.

The electronic versions of the two address registers available were under development, both in terms of the basic information in the registers and how data was presented to ONS. This was an important distinction. For example, the electronic version of the NLPG supplied had an incomplete record of the addresses classified as residential. More complete classification information may have been available within the wider NLPG system, but was not provided to ONS.

This work highlighted important limitations in both products as supplied at the time of research. Key findings are that at that time, both registers were, in varying degrees, deficient in terms of:

- coverage
- classification of residential and non-residential addresses
- positional accuracy of grid references assigned to individual establishments

The figure below shows the key findings in terms of coverage for the address registers.

<i>Census Test area</i>	<i>No. of residential addresses in original list¹</i>	<i>No. of new addresses found in the field</i>	<i>No. of residential addresses in NLPG²</i>	<i>Coverage %</i>	<i>No. of residential addresses in AL2³</i>	<i>Coverage %</i>
Bath and North East Somerset	10,045	920	10,209	93.1	10,321	94.1
Camden	26,355	7,544	32,548	96.0	30,316	89.4
Carmarthenshire ⁴	9,589	385	8,392	84.1	9,616	96.4
Liverpool	41,465	2,523	44,262	100.6	41,520	94.4
Stoke-on-Trent	15,623	343	15,627	97.9	15,741	98.6
Totals	103,077	11,715	111,038	96.7	107,514	93.7

¹ Address list produced from AL2 and used for the address-checking exercise.

² Number of residential addresses identified in the October 2006 supply of NLPG.

³ Number of residential addresses identified in the October 2006 supply of AL2.

⁴ The results for Carmarthenshire were affected by language issues as many addresses in NLPG were recorded in Welsh.

Based on a numerical comparison with the address-checking exercise, the NLPG is shown as having a higher level of coverage within the Test areas with 96.7 per cent compared to 93.7 per cent for AL2.

This may not however, represent the national coverage of the address registers, as the areas known to be harder to enumerate were over represented in the Census Test areas.

The quality of value-added information (namely, building classifications and positional accuracy of grid references) attached to individual addresses for both products as supplied at the time of the research was deficient. The failure to classify many buildings as residential, particularly for the NLPG, may have contributed significantly to their lower than required coverage results. The classification data supplied by Intelligent Addressing were interim classifications, purely to assist ONS research.

Key recommendations and conclusions based on the data supplied are:

- coverage and classification accuracy of the electronic versions of both address registers in the test areas were short of meeting Census Household Frame requirements;
- neither product could have been used to support enumeration activities without major further remedial fieldwork by ONS
- coverage of multiple occupancy addresses was particularly deficient and further work on identifying sub premises is needed
- an alternative solution to implement a 'composite' of the two address registers, taking the best elements of each product, is unlikely to fully identify all addresses

Any future recommendations on the use of address registers for Census 2011, will also consider factors such as licensing costs and terms.

Further research as described in the suppliers' annex, is underway to assess the improvements which are being made by the suppliers. It is also examining other aspects of address registers (for example, coverage of communal establishments and alternative private residence types, identification of areas of change in short periods of time prior to the Census).