UNIVERSITY OF Southampton

Evaluation of automated maintenance procedures

Samantha Cockings and Andrew Harfoot

School of Geography
University of Southampton

September 2010

Produced by the Census2011Geog project
(ESRC Census Programme grant RES-348-25-0019)
http://census2011geog.census.ac.uk

Census.ac.uk

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

CONTENTS

Appendices

Acknowledgements

This research was funded by ESRC Census Development Programme award RES-348-25-0019. The authors are grateful to Professor David Martin, Duncan Hornby and various ONS colleagues (particularly Andy Tait, Donna Viney, Ian Coady, Andy Bates, Steve King and Brian Parry) for their support and advice throughout the project.

The following datasets were used in the research: MasterMap Address Layer 2 (March 2008): under licence from ONS; household-level 2001 Census data and postcoded mid-year population estimates (mid-2007): access granted by ONS MicroData Release Panel, data accessed as Approved Researchers under secure setting conditions at ONS Titchfield; MasterMap Integrated Transport Network layer (December 2007), MasterMap Topographic layer (December 2007), Ordnance Survey Meridian 2 (railways) (October 2008), 2001 Output Area, Lower Layer Super Output Area, Middle Layer Super Output Area boundaries, National Statistics Postcode Directory (February 2008): all obtained from UKBORDERS under academic licence; Universities UK Student Residences List (March 2009) obtained from www.universitiesuk.ac.uk.

Executive summary

This report details the final stage of the ESRC-funded Census2011Geog project, namely the testing and evaluation of the prototype automated maintenance procedures developed by the project. In summary, it concludes that:

1. Automated zone design methods can be used to automatically maintain the 2001 Census output geographies in order to create the output geographies for 2011. The AZTool software provides the functionality required to automatically split, merge or redesign areas according to design criteria specified by ONS.

2. Some areas which had breached thresholds by 2007 had already breached the same thresholds in 2001, and for the same reason in both years. Likewise, some areas would not have breached the upper threshold(s) in 2007 were it not for the presence of a new communal establishment (CE) which did not exist in 2001) or a CE which had grown in population size since 2001. Arguably, areas exhibiting these two types of breaches do not need to be maintained in 2011. ONS needs to decide which categories of threshold breaches to maintain in 2011. These decisions will influence the total number of areas requiring maintenance nationally.

3. Exploration of the spatial distribution of the threshold breaches reveals that in many of the study areas, the over-threshold breaches at the OA level tend to be concentrated in groups of neighbouring OAs, whereas the under-threshold breaches tend to be in isolated OAs. In Southampton and Liverpool, the population growth is mainly in waterfront areas which have been redeveloped or developed from new. In Manchester, the growth is in the city centre, although it is worth noting that this is almost certainly conflated with the under enumeration issue which occurred at the 2001 Census, which makes it more difficult to accurately estimate population change since 2001.

4. There is little difference between the outcomes of adopting a bottom-up (OA-LSOA-MSOA) or a top-down (MSOA-LSOA-OA) approach to the maintenance procedures. Differences are only encountered where an under-threshold geography lies within an over-threshold higher-level geography (e.g. an under-threshold OA within an over-threshold LSOA). Only one such case was encountered in the test data employed for this project. It is suggested that it is more appropriate to adopt a bottom-up approach in 2011 (i.e. maintain OAs first, then LSOAs and then MSOAs) as this ensures that, where possible, all under-threshold OAs are merged to become within threshold, and that this is not influenced by any maintenance carried out at the higher levels of geography.

5. It is recommended that the maintenance is best carried out as an iterative process. The procedures should first be run with all constraints in place. If areas cannot be resolved, the procedures can be re-run, sequentially relaxing constraints in a pre-specified order until all areas are resolved or no further constraints can be relaxed. For the test areas, running the procedures with all constraints meant that a large number of over- and under-threshold areas could not be resolved. Relaxing the minimum boundary length and/or target tolerance constraints enabled substantially more areas to be resolved. In general, it was easier to resolve under-threshold areas (i.e. via mergers) than over-threshold areas (via splits). For some areas, it was possible to quantify the reason why they could not be resolved and to suggest potential semi-automated solutions; for other areas, there was no easily quantifiable reason. ONS needs to decide what to do with the areas which can not be resolved (even after relaxing all of the permitted constraints): they may be left out-with the thresholds, resolved by relaxing further rules or resolved by manual intervention.

6. Using postcodes as the building blocks for the maintenance procedures enabled substantially more over-threshold OAs to be resolved than using street blocks. Using postcodes also resulted in a lower standard deviation around the (household) target but using street blocks lead to more compact maintained output geographies. Surprisingly, there was little difference between the certainty with which postcoded data could be matched to maintained geographies derived from postcode-based versus street block-based maintained geographies. On balance, therefore, it is recommended that ONS employs postcodes as the building blocks for the 2011 maintenance procedures as they enable more over-threshold areas to be resolved and produce OAs which are more internally homogeneous in terms of household size, both of which are key requirements for the 2011 output geographies.

7. In 2001, 98.7% of postcodes within the six study areas matched to one unique OA; by 2007 this percentage had fallen to 93.7%, and, presumably, this percentage will have fallen still further by 2011. Nonetheless, these statistics suggest that, in the vast majority of areas, users should still be able to uniquely match postcoded data to one OA. Of course, problems with matching may be worse in specific areas if the one-to-many relationships are geographically concentrated.

8. The decline in the one-to-one match between postcodes and OAs raises concerns over the confidence users will be able to place in postcode to OA look-ups based on whole- postcode allocation methods (such as the point-in-polygon technique currently employed by the NSPD) in 2011. Evaluation of the differences between the populations assigned from postcodes to OAs via a whole-postcode allocation methodology (whereby the entire postcode's population is redistributed based on the postcode-OA look-ups provided by ONS in 2001 and by the National Statistics Postcode Directory (NSPD) in 2007) compared to a best-fit methodology (whereby a postcode's population is redistributed to OAs based on the individual locations of all of its constituent dwelling spaces) revealed that there had been a substantial reduction in the accuracy of the whole-postcode allocation methodology (compared to the best-fit methodology) between 2001 and 2007. These results raise concerns about the usefulness of such whole-postcode allocation methods for linking and/or re-allocating postcode and census data by 2011.

9. It is recommended that the requirement to keep all parts of a split postcode within the same OA is dropped in 2011 as it places a significant constraint on the ability of the maintenance procedures to find solutions and tends to reduce the statistical and aesthetic attractiveness of any solutions found. As the one-to-one match between postcodes and OAs will have declined in many (un-maintained and maintained) areas anyway, it is no longer considered an essential design requirement for 2011.

10. This research has further demonstrated the importance of having complete and accurate address registers and postcode directories for informing the collection and publication of census small area statistics. The under-enumeration problems experienced in 2001 in Manchester and Westminster arguably resulted in sub-optimal output geographies in this area, which will have since been further compounded by population change. It is suggested that, in areas where the problems are significant, a complete re-design of the output geographies in these areas may be appropriate for 2011; however, users may prefer to keep the existing geographies (except for those areas which have genuinely seen significant population change since 2001) to maximise stability between 2001 and 2011. Whichever way, the automated maintenance procedures developed in this project can be employed, either to completely re-design the areas or to maintain those which breach specified thresholds.

Decisions required by ONS and recommendations from this research

This research has developed and tested methods which can be employed to maintain the 2001 output geographies in order to create the 2011 geographies, using automated maintenance procedures. The implementation of the procedures and final decisions related to the specific design criteria and rule sets to be employed in 2001 now rest with ONS. The key decisions required by ONS, and (in blue) the associated recommendations from this research, are:

1) What approach should be employed for the automated maintenance procedures e.g. bottom-up (OA-LSOA-MSOA) or top-down (MSOA-LSOA-OA)? Bottom-up (OA-LSOA-MSOA)

2) Should any LADs (or other geographical areas) be completely re-designed, rather than just maintaining their constituent zones which have breached thresholds? E.g. Manchester or Westminster? Consider completely re-designing Manchester and Westminster due to 2001 enumeration problems, but consult with users first regarding preferences for stability

3) Should any building blocks be sub-divided prior to carrying out the maintenance procedures?
   a) What measure(s) and threshold(s) should be employed to identify building blocks as candidates for sub-division? Total households (residential only) in building block exceeds upper OA household threshold OR total population (residential + CE) in building block exceeds upper OA population threshold
   b) What methods should be used to sub-divide them? Sub-divide using existing dwelling space grid references or sub-divide manually where this is not possible. Precise methods to be determined by ONS – could be similar to 2001 tower block methods (i.e. move top section of dwelling spaces to nearby location, ensuring both parts of the split tower block are above a specified threshold)

4) Identification of zones requiring maintenance
   a) Thresholds (for each level of output geography)
      i) Should both upper and lower thresholds be employed? Yes
      ii) Should both population and household thresholds be employed? Yes
      iii) What should the threshold values be? See Table 1 in this report
   b) Treatment of CEs
      i) Should CEs not contribute to the household count but contribute their full population to the population count? If so, only the residential dwelling count needs to be tested against the household threshold(s), but residential and CE population counts should be summed to give the total population which needs to be tested against the population threshold(s). CEs do not contribute to household count but do contribute their full population to population total. Households threshold relevant only to residential dwelling count; population thresholds relate to sum of residential and CE populations
   c) Categories of breaches
      i) Should all zones breaching any threshold be maintained, or can some be allowed to remain out-with the thresholds? E.g. zones which already exceeded the upper threshold(s) in 2001 (as there was not an upper threshold then); zones which would not have breached the threshold(s) if only their residential population were tested against the threshold, rather than the residential and CE populations combined (relates to 4b(i) above) In the first instance, try to resolve all zones breaching any thresholds

5) Design criteria and run parameters
   a) Thresholds
      i) Use both upper and lower thresholds? Yes
      ii) Use thresholds for both total households and total population? Yes
      iii) Values for each of 5a(i) and 5a(ii) for each output geography level? See Table 1 in this report
   b) Target
      i) Use household target or population target? Household target
      ii) Value of 5b(i) for each output geography level? See Table 2 in this report
      iii) Target tolerance for initial random aggregation (IRA) (if applicable)? N/A
   c) Homogeneity
      i) Use or not? Use
      ii) What score/measure? Intra-area correlation (IAC) score
      iii) What variables and categories? Tenure and Accommodation type. Categories to be confirmed by ONS, depending on 2011 Census questions/categories[a]. Suggestions: *Tenure*: owns; shared; rents; rfree. *Accom type*: detached; semi; terraced; flat; parthouse; commerce; nonperm.
      iv) Weighting for each variable? 100
   d) Shape constraint
      i) Use or not? Use
      ii) What score/measure? Compactness[b] = Perimeter^2/Area
      iii) Weighting? 100
   e) Minimum boundary length
      i) Use or not? Use
      ii) What value? 10%

[a] The categories for accommodation type in 2011 will be the same as those used in 2001 and in our research, whereas the 2011 categories for tenure will change slightly. ONS will need to consider which categories/groupings to use: the aim is to have categories which capture socio-economic variation between areas. Suggestions are provided below:

2011 Accommodation type codes and suggested categories

Suggested category	*2011 Code*	*2011 Description*
Detached	1	Detached
Semi	2	Semi-detached
Terraced	3	Terraced (including end-terrace)
Flat	4	In a purpose-built block of flats or tenement
Parthouse	5	Part of a converted or shared house (including bed-sits)
Commerce	6	In a commercial building (for example, in an office building, hotel, or over a shop)
Nonperm	7	A caravan or other mobile or temporary structure

2011 Tenure codes and suggested categories

Suggested category	*2011 Code*	*2011 Description*
Owns	1	Owns outright
Owns	2	Owns with a mortgage or loan
Shared	3	Part owns and part rents (shared ownership)
Rents	4	Rents (with or without housing benefit)
Rfree	5	Lives here rent free

[b] AZTool and the research reported here employ perimeter^2/area as the shape constraint. This is different to the shape metric used in OAPS in 2001, which was based on the distribution of addresses/postcodes: ONS will need to decide which shape metric to use. Different shape metrics can be programmed and implemented in AZTool if required.

f) Regional constraint
   i) Use or not? Do not use (higher level geography constraint is imposed by only supplying AZTool with zones within the higher level geography e.g. candidate OAs within an LSOA when merging, or postcode building blocks within an individual OA when splitting, therefore no need to use the regional constraint option provided in AZTool)
   ii) What higher level output geography to use at each level? N/A
g) Use IRA method which force number of tracts to equal total <target variable> divided by <target>, or method which uses target tolerance? Use tracts = <target variable> divided by <target> method
h) Number of iterations for IRA? Cannot be predetermined – requires some experimentation at the start of the process. For the research reported here, 5 iterations were used.
i) Number of runs (for full automated zoning procedure (AZP))? Again, needs experimentation at start of process. 100 runs were used in this research. More runs are likely to be needed for any areas which require complete re-design, compared to areas where small sub-sets of zones are being split or merged.
j) Allow donuts in output tracts or not? Allow donuts
k) Ignore cases of Bishop's contiguity or not (when determining neighbours)? Ignore

6) Ruleset for relaxation of design criteria
   a) What design criteria can be relaxed, how, and in what order? Depends on priorities set by ONS. In this research, we relaxed (i) minimum boundary length constraint, (ii) target tolerance, (iii) both minimum boundary length and target tolerance.

7) What to do with unresolved areas? Depends on priorities set by ONS. Suggest that all under-threshold OAs *must* be resolved manually if they have not been resolved by AZTool. Under-threshold LSOAs and MSOAs *could* be allowed to remain so, provided that they are within a pre-specified tolerance of the lower threshold(s)? Over-threshold zones *could* also be allowed to remain so, as long as they are within a specified tolerance of the upper thresholds? Otherwise, both will probably require manual intervention. Note that any manual solutions will almost certainly need to breach the one-to-one 2001:2011 lower-level:higher-level output geography relationships.

Note that, with the exception of 1 (above), these decisions do not include those related to the creation of building blocks for the automated maintenance procedures: these methods and associated decisions have already been outlined in previous papers to OAMPAG.

List of Tables

List of Figures

Evaluation of automated maintenance procedures

1 Background

Earlier reports to OAMPAG presented the results of an analysis of population change between 2001 and 2005/06. These analyses were then used to select six study areas: Camden, Isle of Anglesey (hereafter referred to as Anglesey), Lancaster, Liverpool, Manchester and Southampton. Test data for 2007/8 (hereafter referred to as 2007) were then created for the six study areas using methods detailed in previous OAMPAG papers. These data provided estimates of the population, tenure and accommodation type counts at household level for the study areas. Two sets of building blocks were then created for each study area, one based on postcodes and one on street-blocks. Algorithms for the automated maintenance procedures were coded, according to specifications previously approved by OAMPAG. The automated maintenance software is called AZTool, and is written in Visual Basic .NET. It employs a very similar algorithm to that used by the OAPS process in 2001. This report details the final stage of the ESRC-funded Census2011Geog project, namely the testing and evaluation of the automated maintenance procedures, using the test data within the six study areas.

2 Aims

The aims of the Census2011Geog project were:

1. To develop and evaluate automated procedures for maintaining (splitting, merging, re-designing) the 2001 census output geographies in order to create 2011 output geographies.
2. To assess the implications (in terms of the achievement of optimal solutions, the statistical characteristics of resultant zones, and the effects on data linkage) of using different building blocks (such as postcodes and street blocks) for these maintenance procedures.

3 Methods

The 2007 test data were aggregated to the 2001 output geographies and those areas breaching the thresholds shown in Table 1 identified.

Table 1 Thresholds

	Population thresholds		***Household thresholds***	
Geography	**Lower**	**Upper**	**Lower**	**Upper**
OA	100	625	40	250
LSOA	1,000	3,000	400	1,200
MSOA	5,000	15,000	2,000	6,000

Reasons for the breaches were explored, including which particular thresholds (i.e. population and/or household) they had breached, whether or not they had already breached thresholds in 2001, and whether the breaches would not have occurred were it not for the presence of a CE population (in 2001 and/or 2007). As outlined in previous OAMPAGs, maintenance of these geographies can either be carried out using a bottom-up approach (i.e. maintain OAs first, followed by LSOAs and then MSOAs), or a top-down approach (MSOAs-LSOAs-OAs). AZTool was employed to

carry out the maintenance procedures using both approaches, using the constraints and criteria shown in Table 2.

Table 2 Constraints and criteria employed in the maintenance procedures

Constraint/criteria	*Details*	*Weighting*
Thresholds	As per Table 1	N/A
Target (households)	OA: 125; LSOA: 600; MSOA: 3,000	100
Homogeneity	Intra-area correlation scores for accommodation type and tenure	100
Shape	Perimeter^2/Area	100
Minimum boundary length	10% of the total perimeter	N/A
Regional constraint	Respect higher-level output geographies	N/A

For areas where solutions could not be found using all of these constraints/criteria, an iterative process of relaxing constraints and re-running the procedures was undertaken. First, the minimum boundary length constraint was relaxed; then the target tolerance; and finally both were relaxed together. Any areas for which solutions had not been found were left unresolved. Where there was an identifiable and quantifiable reason for this non-resolution, this was recorded.

When splitting OAs, two different sets of building blocks were employed: postcodes and street blocks. The maintained geographies for the top-down and bottom-up approaches, using both the postcodes and street blocks, were then re-combined to produce four sets of maintained output geographies for all study areas (i.e. MSOAs, LSOAs and two sets of OAs (one from postcodes and one from street blocks). The statistical characteristics of these maintained geographies were then recorded and compared with those of the 2001 and 2007 geographies. The differences between the top-down and bottom-up approaches and the impact of using postcodes or street blocks were evaluated.

4 Results and discussion

4.1 Breaches of the output geographies by 2007

Table 3 shows the number of threshold breaches in 2007 for all six study areas combined, classified by type of breach, and by output geography level. The over-threshold breach category is further sub-divided to identify situations where there is a weaker case for maintenance. These are: (i) those areas which exceeded the upper population threshold in 2007, but which would not have exceeded the threshold when considering their residential population alone (i.e. a CE population was required to take it over the upper threshold) ***and*** where this was also the case in 2001; (ii) the same scenario as in (i), but where this had ***not*** been the case in 2001 i.e. areas which had seen a new CE, or their CE population had grown, or their residential population had decreased, since 2001; (iii) those which exceeded one or both of the upper (i.e. population or household) thresholds in 2007, but which were already over the same threshold(s) for the same reason in 2001 (but excluding the cases described in (i)); and (iv) those which exceeded one or both of the upper thresholds, but not in the manner described in (i), (ii) or (iii). The same results, broken down by study area, are presented in Appendix A.

Table 3 Threshold breaches in 2007, classified by type of breach and reason for over-threshold breaches, for all study areas combined, by output geography level

	Total	Under threshold	Within threshold	Over threshold			
				As at 2001			Other reasons
					Resid pop below upper threshold		
OAs	4988	43	4836	2	18	10	79
LSOAs	962	12	938	0	2	1	9
MSOAs	200	1	198	0	0	0	1
Case				*(iii)*	*(i)*	*(ii)*	*(iv)*

It is clear that the vast majority of zones (at all levels) were within-threshold in 2007. The within-threshold areas clearly do not need to be maintained. It is reasonable to assume that the areas which are under-threshold and those failing the upper threshold tests for an unspecified reason (i.e. case (iv) above) definitely need to be maintained. It is debatable though, whether the other categories (i.e. cases (i), (ii) and (iii) above), should be maintained or not. One could argue that those which were already breaching thresholds in 2001 and which were breaching the same thresholds in 2007 (cases (i) and (iii)) need not be maintained. Likewise, it would be defensible to argue that the presence of a large CE within an area ((i) and (ii)) is insufficient grounds for maintaining an area. These are policy and statistical disclosure control (SDC)-related decisions which will need to be taken by ONS ahead of the 2011 Census. Clearly, such decisions may substantially alter the total number of areas requiring maintenance across England and Wales, particularly at the OA level. In the study areas here, the minimum numbers of areas requiring maintenance are 122 OAs, 21 LSOAs and 2 MSOAs, compared to a maximum of 152, 24 and 2 respectively if all breached areas are maintained. For the purposes of the research reported here, it was assumed that ***all*** under- and over-threshold output geographies should be maintained, even if they were already breached in 2001 or if the reason for their breach in 2007 was due to the presence of a CE population. This therefore represents the worst-case scenario, where all breaches have to be dealt with.

Geographically, there are differences between the areas experiencing population growth and population decline. In some of the study areas, population growth is clustered in distinct groups of neighbouring OAs, for example along waterfront locations in Southampton and Liverpool, and in the city centre of Manchester. It should be noted though that at least some of the areas exhibiting apparent growth in Manchester city centre were almost certainly also some of the same areas which experienced under enumeration in the 2001 Census, which will have distorted the accuracy of any change statistics calculated for this area. By contrast, in Camden and Lancaster, growth is found in individual, isolated, OAs. Some of these are due to the presence of special populations, such as student halls of residences. In all of the study areas, the declining populations tend to be in individual, isolated, OAs. There are no particular spatial trends worthy of note in Anglesey, which is not surprising given that it was included as a "control" study area, with a relatively stable population since 2001.

4.2 Evaluation of bottom-up and top-down approaches to maintenance

The maintenance processes may be carried out using either a bottom-up (i.e. maintain OAs first, then LSOAs and then MSOAs) or top-down (MSOAs-LSOAs-OAs) approach. Both processes produce very similar results in terms of the maintained geographies. Differences potentially arise when an under-threshold geography (e.g. an OA) sits within an over-threshold higher-level geography (e.g. an LSOA). In this situation, the order in which the maintenance is carried out influences either the

ability to fix the higher-level geography (in the bottom-up approach) or the ability to fix the lower-level geography (in the top-down approach). There was only one case of this in the entire six study areas. Whilst it is impossible to predict the number of times that this situation may occur nationally in 2011, the six study areas selected here (other than Anglesey) were chosen because they had experienced significant population change since 2001 and were thought to be representative of the levels and types of changes likely to be seen in 2011. It is reasonable to assume, therefore, that there will not be many situations like this arising in 2011. Nonetheless, explicit thought should be given (by ONS) to the preferred approach for 2011. Given that ONS has previously indicated that adherence to the lower thresholds is critical (for SDC reasons), it is suggested that it may be more appropriate to adopt a bottom-up approach in 2011 as this ensures that, where possible, all under-threshold OAs are merged to become within threshold, and that this is not influenced by any maintenance carried out at the higher levels of geography. It should be noted that the consequence of doing this may be that a small number of higher level geographies e.g. LSOAs remain over-threshold, but this is thought to be less critical. For conciseness, the rest of this report presents only the results for the bottom-up approach.

4.3 Maintained geographies: numbers resolved or unresolved

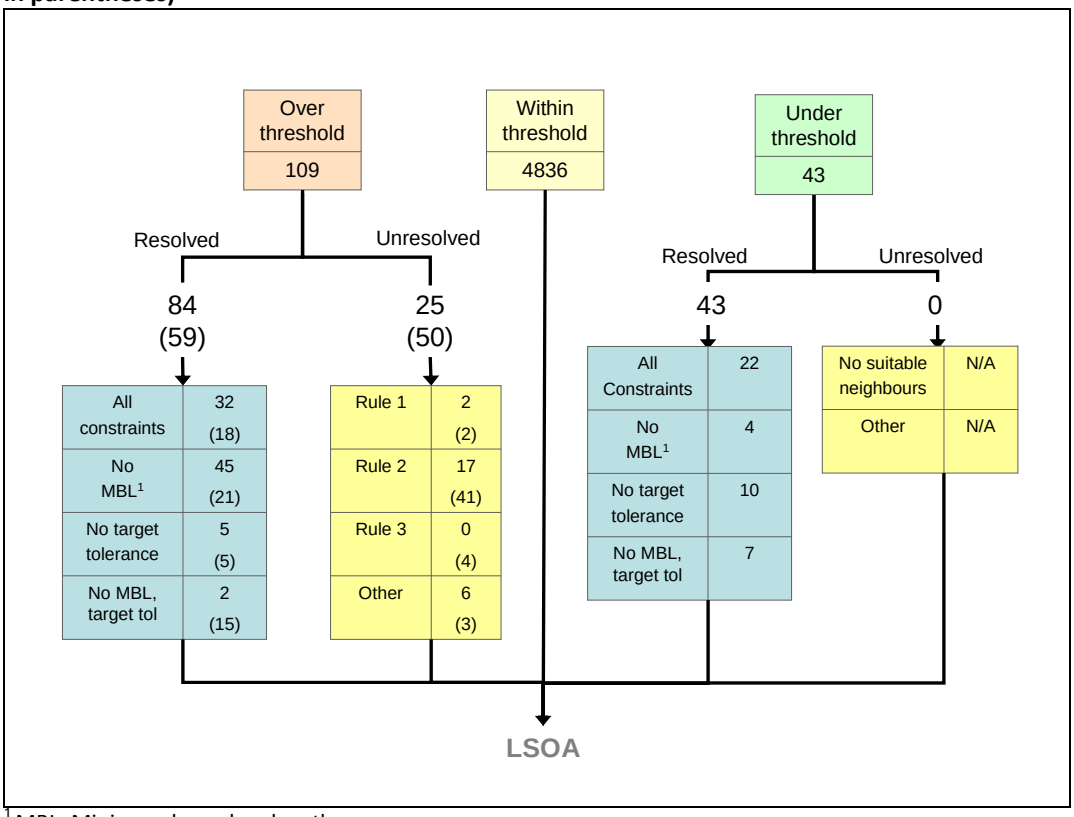
Figures 1 to 3 show how many of the over- and under-threshold output geographies were resolved or unresolved at each level for all study areas using the bottom-up approach. Appendix B (Tables B1-B4) shows the same statistics in tabular format for each individual study area. For the resolved areas, the schematics show how many were resolved using all constraints and how many were subsequently resolved by sequentially relaxing the minimum boundary length constraint, the target tolerance and finally both the minimum boundary length and the target tolerance. Note that the effect of relaxing the minimum boundary length is a reduction in the compactness of shape of the maintained output geographies; relaxing the target tolerance potentially reduces their homogeneity of population size.

For those over-thresholds areas which could not be resolved, the reasons for their non-resolution were categorised as follows:

- **Rule 1**: Area has breached upper population threshold but has less than two times the lower household threshold (or vice versa) so therefore cannot be split into new within-threshold areas.
- **Rule 2**: Area has breached upper population and/or household threshold(s) but one or more of its constituent building blocks has also breached the same upper threshold so it cannot be split to produce areas which are all within-threshold.
- **Rule 3**: Area has breached upper population and/or household threshold(s) but population and/or household distribution of constituent building blocks is overly concentrated within one building block, leaving insufficient population and/or households in the other building blocks to create within threshold areas.

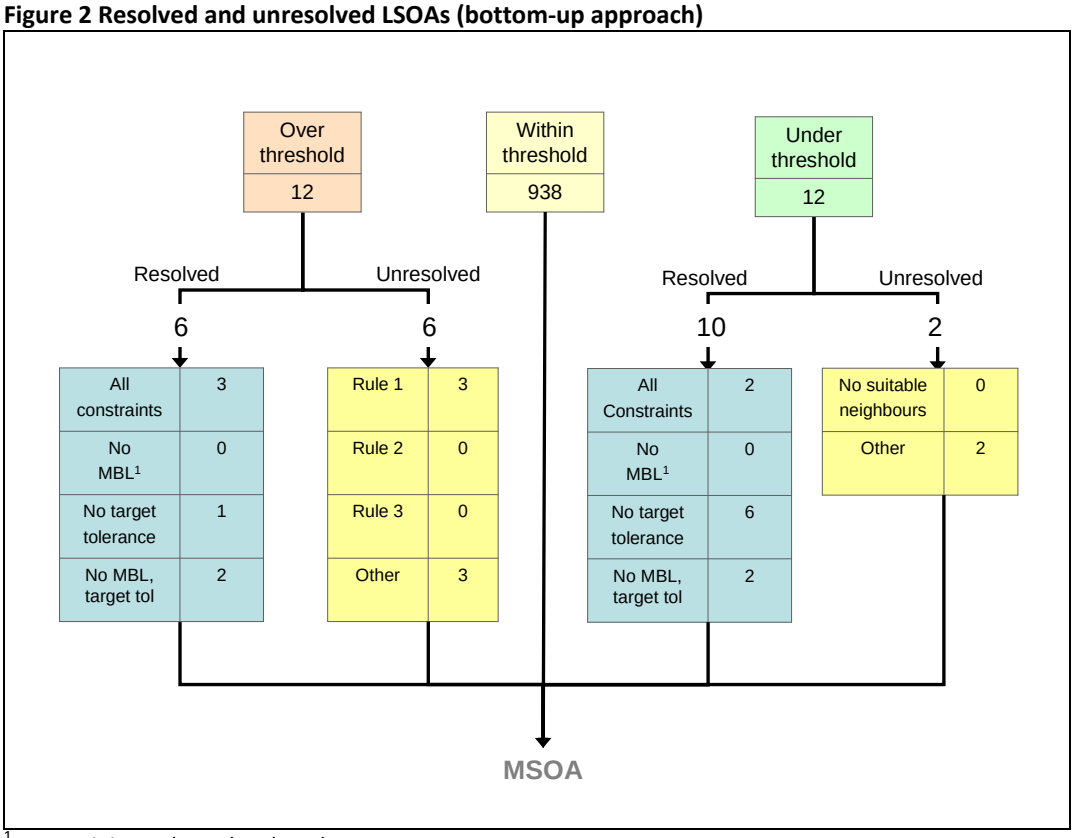
Over-threshold areas which were not resolved but whose reason for non-resolution could not be quantitatively classified by one of the rules above, were recorded in the category “Other”. For the under-thresholds areas, those which could not be resolved due to a lack of suitable neighbouring building blocks with which to merge were categorised as “No suitable neighbours”; all other unresolved under-threshold areas were recorded as “Other”.

Figure 1 Resolved and unresolved OAs (bottom-up approach)
(Note: for over-threshold OAs, counts for postcodes are shown on top of counts for street blocks, which are in parentheses)



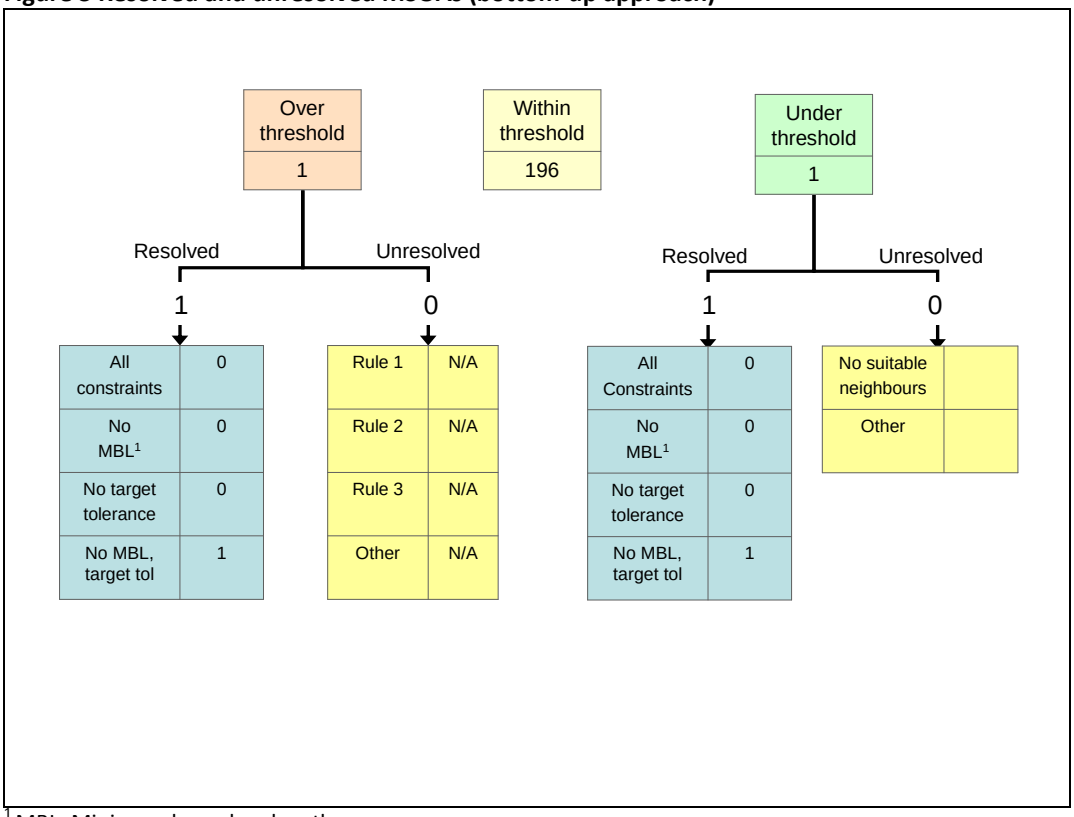
[1] MBL: Minimum boundary length

Figure 2 Resolved and unresolved LSOAs (bottom-up approach)



[1] MBL: Minimum boundary length

Figure 3 Resolved and unresolved MSOAs (bottom-up approach)



[1] MBL: Minimum boundary length

Ideally, we would employ all of the constraints when carrying out the maintenance, but it is clear that doing so results in large numbers of areas remaining unresolved (only 25% of over-threshold LSOAs; 17% of under-threshold LSOAs; 29% and 17% of over-threshold OAs using postcode or street block building blocks respectively; and 51% of under-threshold OAs were resolved when all constraints were in place). Relaxing the minimum boundary length and/or the target tolerance constraints does, however, substantially increase the number of areas which can be resolved. By relaxing one or both of these constraints, we can resolve 50% of over-threshold LSOAs; 83% of under-threshold LSOAs; 77% and 54% of over-threshold OAs using postcode or street block building blocks respectively; and 100% of under-threshold OAs. The two MSOAs requiring maintenance (one over- and one under-threshold) were both resolved by relaxing the minimum boundary length and target tolerance constraints.

The results show that it was generally easier to resolve under-threshold areas (i.e. mergers) than over-threshold areas (i.e. splits) and that postcode building blocks afforded the capability to resolve substantially more over-threshold OAs than street blocks.

At the OA level, the main reason for not being able to resolve over-threshold OAs was due to individual building blocks within the OA having populations or household counts which were themselves greater than the OA-level upper threshold(s). This was particularly problematic when using the street block building blocks, where 41 (out of 109) over-threshold OAs could not be resolved for this reason, compared to 17 for the postcode building blocks. Figure 4 presents an example of this in Camden, where a street block building block within an over-threshold OA has a population (762) which is higher than the OA-level upper threshold (625). It is therefore not possible to sub-divide the OA in order to create within-threshold tracts using the street block building blocks. By contrast, the postcode building blocks layer for the same over-threshold OA

contains six postcodes (also shown in Figure 4) within the problematic street block, all of which have lower and more equal populations (totals not shown for disclosure control reasons): this allows the successful sub-division of the OA into within-threshold tracts. Interestingly, whilst we might expect the reverse to be true i.e. situations where the street blocks enable a solution to be found but the postcodes do not, no such situations were encountered in the study areas explored here. Arguably, this may be a reflection of the types of areas which experience change, and the very close relationship between the introduction and recycling of postcodes in areas of population/housing development. Whilst street blocks clearly also reflect change e.g. with new roads being built in new housing developments, they do not tend to sub-divide areas of high density population (such as flats, student residences, high-density housing estates, etc.) as well as postcodes do. As population growth since 2001 has often tended to occur in these types of areas, in this respect postcodes may provide a more suitable building block base than the street blocks. In addition, and as discussed in more detail later, the average population and household size of the street blocks is almost twice the size of the postcodes, and some street blocks have extremely high populations e.g. the highest is a student residence in Lancaster, with 4278 people! The postcodes therefore provide more granularity and flexibility for the maintenance procedures.

Figure 4 Example of Rule 2: non-resolution of over-threshold OA in Camden using street blocks (One of constituent street blocks already breaches OA's upper threshold; solution is, however, possible using postcodes, which have smaller and more evenly distributed populations)

By contrast, at the LSOA level, the main reasons for non-resolution of the over-thresholds areas were insufficient household counts to enable the LSOA to be split (3 of the 6 unresolved LSOAs) or a geometric configuration which prevented a solution being found (2 out of 6). Interestingly though, all three of the LSOAs which could not be split due to insufficient household counts, would not have breached the upper population threshold on the basis of their residential population alone (i.e. excluding their CE population), and two of these would have already exceeded the upper population threshold in 2001 had one been in place at that time. Arguably then, these two LSOAs need not have been maintained anyway.

4.4 What to do with the unresolved areas?

Where it has not been possible to find a solution for an over- or under-threshold area, ONS needs to decide whether to insist that the area be maintained (in which case, other rules will need to be relaxed) or whether there are situations were an over- or under-threshold area will be allowed to remain so.

Presumably, under-threshold areas (especially OAs) will need to be resolved to ensure that they meet statistical disclosure controls. In these instances, assuming that merging with a part-area (e.g. merging an LSOA with an OA) is not an option due to a need to maintain the existing one-to-one relationships, the only solution would be to merge the under-threshold area with an adjacent over-threshold neighbour: this will produce an over-threshold area, but this is assumed to be preferable to leaving an area under-threshold.

Whether over-threshold areas should be allowed to remain so is more debatable. As mentioned previously, some areas would have already been over-threshold in 2001 had there been an equivalent upper threshold in place then. Arguably then, these areas could be allowed to remain over-threshold in 2011. Or, it could even be argued that attempts should not be made to maintain them at all.

For over-threshold areas which cannot be resolved because they have failed Rule 2 (i.e. because a constituent building block has an over-threshold population and/or household count) and which ONS decides cannot be allowed to remain so, the only real option is to attempt to sub-divide any problematic building blocks prior to maintenance. In 2001, prior to carrying out the OA design process, any set of residential dwelling spaces coincident on one grid-reference which had a combined household count of greater than 250 was classified as a tower block. ONS then manually moved a sub-set of these households/dwelling spaces (grouped by postcode) to one or more nearby locations in order to produce sub-threshold building blocks prior to carrying out the automated design process. ONS needs to decide whether to deal with residential tower blocks in the same way in 2011, and whether to use the same threshold (i.e. 250 households coincident on one grid reference) to identify them. Table 4 presents the number of residential tower blocks in each of the study areas (for postcode-based building blocks), for a range of potential thresholds. Camden clearly has the largest number of tower blocks, followed by Manchester and Southampton. Our analysis suggests that there are five tower blocks having over 250 coincident households in the six study areas. The largest is a block in Camden, which contains 472 households on one grid reference. By contrast, if a threshold of 100 coincident households were used, there would be 76 instances of tower blocks requiring manual intervention.

To the authors' knowledge, no sub-division of CEs with large populations was undertaken in 2001.

Table 4 Cumulative number of tower blocks in each study area, using a range of thresholds for identifying tower blocks

	Cumulative number of tower blocks, by threshold and study area						
Threshold: number of households coincident on one grid reference	Anglesey	Camden	Lancaster	Liverpool	Manchester	Southampton	Total
100	0	36	0	8	18	14	76
150	0	12	0	5	5	2	24
200	0	2	0	4	3	0	9
250	0	2	0	2	1	0	5

Whilst the above analysis is useful and consistent with the 2001 design process, the 2011 maintenance process is concerned not only with identifying and maintaining areas which have fallen below the lower thresholds, but also those which have grown excessively large in terms of either or both of their population and/or household counts. ONS needs to decide whether to limit these checks to just residential households (as in 2001) or whether to include CEs with large populations. Any building blocks which themselves fail either of the upper OA-level population and/or household threshold checks could be identified in advance of carrying out the maintenance process, as they will prevent the successful splitting of the relevant OA. In order to provide an idea of the magnitude of the problem and the types of situations likely to be encountered, Tables 5(a) and 5(b) present a summary of the building blocks in all study areas exceeding the upper OA-level population and/or household thresholds in 2007, for the postcode- and street block-based building blocks respectively. The building blocks are grouped into four categories, related to their potential for sub-division:

(i) the building block can be sub-divided into smaller sub-threshold building blocks using the constituent (dwelling space) grid-references within the building block

(ii) the building block is a CE tower block (i.e. all of the population within the building block is located on one grid-reference and the population count exceeds 625 and the household count is zero) and thus requires manual intervention

(iii) the building block is a residential tower block (i.e. all of the households are located on one grid-reference and the household count exceeds 250) and thus requires manual intervention

(iv) the building block requires manual intervention as it contains parts which cannot be sub-divided into sub-threshold building blocks using its constituent dwelling space grid-references, but it is not a case of (ii) or (iii).

Case (i) (i.e. sub-division based on existing grid-references) would require manual intervention, but this process would be amenable to automation in the form of GIS processing scripts. Cases (ii-iv) would require user-intervention to visualise the specific situation and make decisions about how to sub-divide the block (with case (iv) being the most labour-intensive). These analyses show that 17 out of the 20 over-threshold postcode-based building blocks would require manual intervention, compared to 17 of the 43 of the street blocks, but that all over-threshold postcode building blocks could be resolved using a combination of automatic sub-division based on existing grid-references or manual resolution of tower block situations.

Table 5(a) Postcode-based building blocks exceeding OA-level upper population and/or household threshold(s), classified by potential for sub-division

Case	*Category*	*Count*
(i)	Pop and/or HHD can be sub-divided to give sub-threshold BBs	3
(ii)	CE tower block - requires manual intervention	12
(iii)	Residential tower block - requires manual intervention	5
(iv)	Pop and/or HHD requires manual intervention	0
	Total	20

Table 5(b) Street block-based building blocks exceeding OA-level upper population and/or household threshold(s), classified by potential for sub-division

Case	*Category*	*Count*
(i)	Pop and/or HHD can be sub-divided to give sub-threshold BBs	26
(ii)	CE tower block - requires manual intervention	3
(iii)	Residential tower block - requires manual intervention	2
(iv)	Pop and/or HHD requires manual intervention	12
	Total	43

It may also be possible to resolve those areas remaining over-threshold due to Rule 3 (i.e. an uneven household and/or population distribution within their building blocks) by a similar method of manual sub-division, but this will be dependent on the distribution of population/households within the constituent building blocks. Other unresolved areas are likely to have to be considered on a case-by-case basis.

4.5 Statistical characteristics of the maintained output geographies

Table 6 presents the statistical characteristics of the (bottom-up) maintained OAs using postcodes and street blocks, compared with the characteristics of the 2001 OAs and the 2007 data in 2001 OAs, for all study areas combined together. Note that the statistics for the maintained geographies include areas which were not resolved by the maintenance procedures. These results show that the means and standard deviations of the total population and total households per OA had increased between 2001 and 2007, whereas the homogeneity of tenure and accommodation type within OAs had decreased. Following maintenance, the means and standard deviations of total population and total households for the maintained OA geographies using both postcodes and street blocks moved back closer to the 2001 mean and standard deviation. There was little improvement in the homogeneity levels following the maintenance procedures, presumably because the algorithm prioritises improvement of the distribution of the target variable (households) and ensuring that the thresholds are met. Interestingly, the shape scores for the postcode- and street block-based maintained OAs were actually very slightly better (i.e. more compact) than the original 2001 OAs, with the street blocks, in particular, producing more compact shapes. This may be due to the fact that the maintenance algorithms did not insist that split postcodes were placed within the same OA whereas the original 2001 OAPS process did this and it acted as a significant constraint on the algorithm, thus reducing its ability to produce compact shapes and homogeneous areas.

Tables 7 and 8 present the equivalent results for LSOAs and MSOAs. As with the OAs, there were clear increases in the mean and standard deviation of total population per LSOA and MSOA, and reductions in the homogeneity of tenure and accommodation type within LSOAs and MSOAs, between 2001 and 2007. As with the OAs, the maintenance procedures improved the standard deviations of the households for both the LSOAs and MSOAs compared to their 2007 equivalents, although the mean total number of households per LSOA and MSOA did not really change. In contrast to the OAs, the maintenance procedures had very little impact on the means and standard deviations of the total population per LSOA/MSOA and, in some cases, such as for the standard deviation of total population and the shape scores for the MSOAs, they actually resulted in slightly worse statistics. This is probably because the numbers of breached /maintained areas at the LSOA and MSOA levels are much lower than at the OA level and the number of potential solutions is much smaller, resulting in a greater likelihood of lower quality solutions than if designing from new.

It is notable that fixing under-threshold areas resulted in less improvement in the overall statistical characteristics of the geographies than fixing over-threshold areas, and, in some cases, actually reduced the overall quality of the solution. This is because there were generally fewer under-

threshold areas, the number of candidate (under- and within-threshold) neighbours for merging was lower, and, in many cases, the only option was to merge with a neighbour which was actually already within-threshold: this tended to reduce the overall quality of the solution. In contrast, the over-threshold areas were resolved by splitting the over-threshold zones internally: where a solution could be found, this tended to result in improvements to the overall statistical quality of the geographies. Note though that, as described previously, fewer solutions were found overall for over-threshold areas than under-threshold areas. The maintenance process is also quite strongly focused on achieving the desired (population and household) thresholds and the household target: this meant that whilst the mean and SD of the households in an area were generally improved by the maintenance process, the mean and SD of the population need not have been: this was especially true for the maintenance of under-threshold areas.

The results for the original and maintained geographies within each of the study areas are presented in Appendix C (Tables C1-C3). Overall, the patterns observed in the individual study areas were consistent with those identified for all study areas considered together, although it should be noted that a number of the study areas had very low numbers of zones requiring maintenance so the changes were minimal in these areas. The statistical characteristics of the maintained geographies (at all levels) in Manchester showed clear improvement compared to the values observed in the (deteriorating) geographies in 2007. This was due to the successful maintenance of a large number of over-threshold areas, which had probably arisen due to the under-enumeration problems experienced in this area.

4.6 Implications of using postcodes or street blocks as building blocks

4.6.1 Impact on statistical and aesthetic characteristics of maintained zones

Table 9 presents the statistical characteristics of the postcode- and street block-based building blocks. The street blocks are almost exactly twice the (population and household) size of the postcodes, and are less homogeneous in terms of population and household size than the postcodes. The postcodes are also more homogeneous in terms of tenure and accommodation type, presumably due to their smaller (areal) size, but the street blocks are very slightly more compact. Table 10 shows the percentage of postcodes overlapping with one or more street blocks; Table 11 shows the converse. 74% of postcodes in all study areas overlap with just one street block; by comparison, only 36% of street blocks overlap with only one postcode, illustrating the relative difference in their sizes. This is not surprising, given the characteristics observed above. The relationship between postcodes and street blocks varies by study area: Camden and Southampton exhibit the closest one-to-one geographic match between postcodes and street blocks, whereas in Liverpool, Lancaster and Anglesey more postcodes overlap with a greater number of street blocks. The intersection of street blocks with postcodes is noticeably different in Anglesey compared to the other study areas: more street blocks overlap with a greater number of postcodes, presumably due to its predominantly rural nature and hence large sprawling postcodes.

Using postcodes as the building blocks for the maintenance procedures enables more over-threshold OAs to be resolved than when using street blocks. The postcodes also result in a lower standard deviation around the (household) target. This is because the postcodes are smaller in areal extent and in population and household size than the street blocks. By contrast, the street blocks facilitate the production of more compact maintained geographies, presumably because they are more compact themselves. Figure 5 provides an example of this, for an over-threshold OA in Southampton, where the solution produced by the street block building blocks is more aesthetically pleasing than that produced using postcode building blocks. The shapes of the newly maintained

street block-based OAs are more compact and the boundaries lie along roads, whereas the postcode-based OAs are more angular and cut across geographical features. But, it is important to note that, overall, the statistical qualities of the postcode-based maintained OAs are better than those of the street block-based OAs. This is a good example of the classic trade-off between the aesthetic and statistical characteristics of zoning systems.

Table 6 Statistical characteristics of maintained OAs (using postcodes and street blocks), compared to 2001 and 2007 data in 2001 OAs, for all study areas

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	4988	290.4	101.9	124.8	16.3	0.182	0.289	37.83	14.44
2007	4988	314.6	140.7	127.7	44.0	0.161	0.263	37.83	14.44
Maintained (postcodes)	5074	309.3	128.6	125.5	29.5	0.162	0.264	37.79	14.47
Maintained (street blocks)	5021	312.5	133.3	126.8	36.8	0.161	0.264	37.70	14.37

Table 7 Statistical characteristics of maintained LSOAs compared to 2001 and 2007/08 data in 2001 LSOAs, for all study areas

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	962	1505.7	201.7	646.9	101.6	0.132	0.190	42.70	15.02
2007/08	962	1631.2	362.7	662.0	171.7	0.117	0.177	42.70	15.02
Maintained	961	1632.9	321.1	662.7	132.3	0.117	0.177	42.74	14.96

Table 8 Statistical characteristics of maintained MSOAs compared to 2001 and 2007/08 data in 2001 MSOAs, for all study areas

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	200	7242.5	1078.9	3111.7	472.5	0.091	0.134	44.42	15.15
2007/08	200	7846.3	1465.0	3184.1	614.5	0.083	0.129	44.42	15.15
Maintained	200	7846.3	1535.4	3184.1	588.3	0.084	0.128	44.60	15.23

Table 9 Statistical characteristics of postcode- and street block-based building blocks

		Total population			Total households			Homogeneity (IAC)		Compactness (P^2/A)	
	Count	Mean	Range	SD	Mean	Range	SD	Tenure	Accommodation type	Mean	SD
Postcodes	43211	36.3	0-1206*	40.9	14.8	0-472	15.0	0.239	0.479	25.36	10.37
Street blocks	21627	72.6	0-4278*	86.2	29.4	0-605	31.3	0.196	0.364	27.07	11.57

* 424 Street blocks and 1575 Postcode building blocks have zero population: these are building blocks which contain only vacant dwelling spaces, which were included in the creation of the building blocks (as per the 2001 OAPS process)

Table 10 Percentage of postcodes overlapping with given number of street blocks

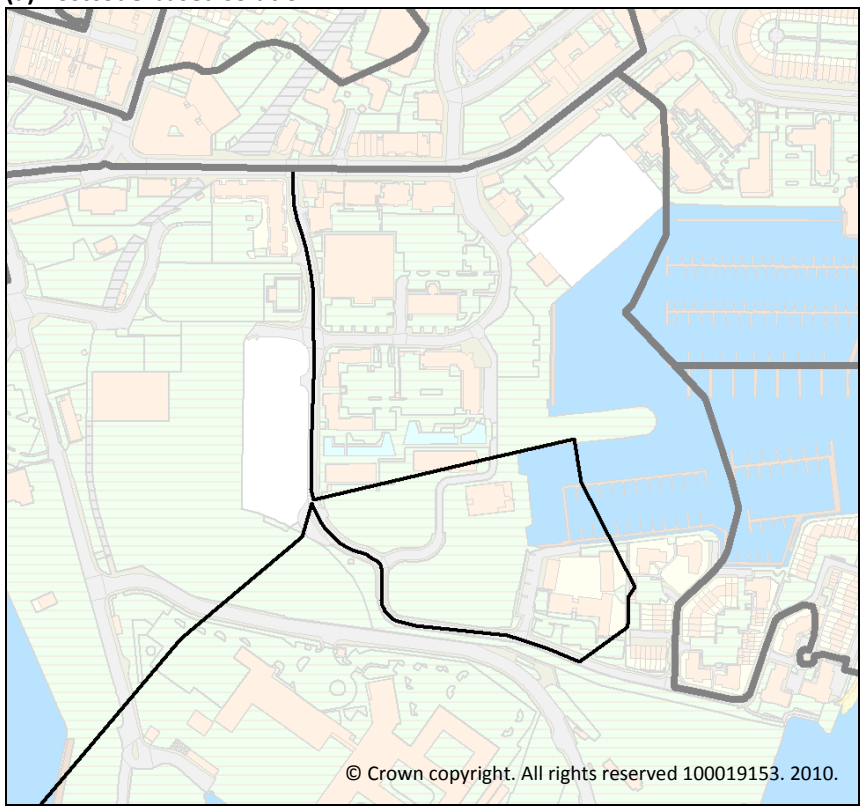
Number of street blocks overlapped with	All	Anglesey	Camden	Lancaster	Liverpool	Manchester	Southampton
1	74.0	60.9	89.8	62.2	66.8	74.5	88.2
2	20.8	31.2	9.5	29.8	24.7	20.8	10.9
3	3.9	5.9	0.7	6.1	6.0	3.7	0.7
4	1.0	1.6	0.0	1.6	1.6	0.8	0.1
5	0.3	0.3	0.0	0.3	0.7	0.2	0.0
6-10	0.1	0.1	0.0	0.0	0.2	0.0	0.0
> 10	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 11 Percentage of street blocks overlapping with given number of postcodes

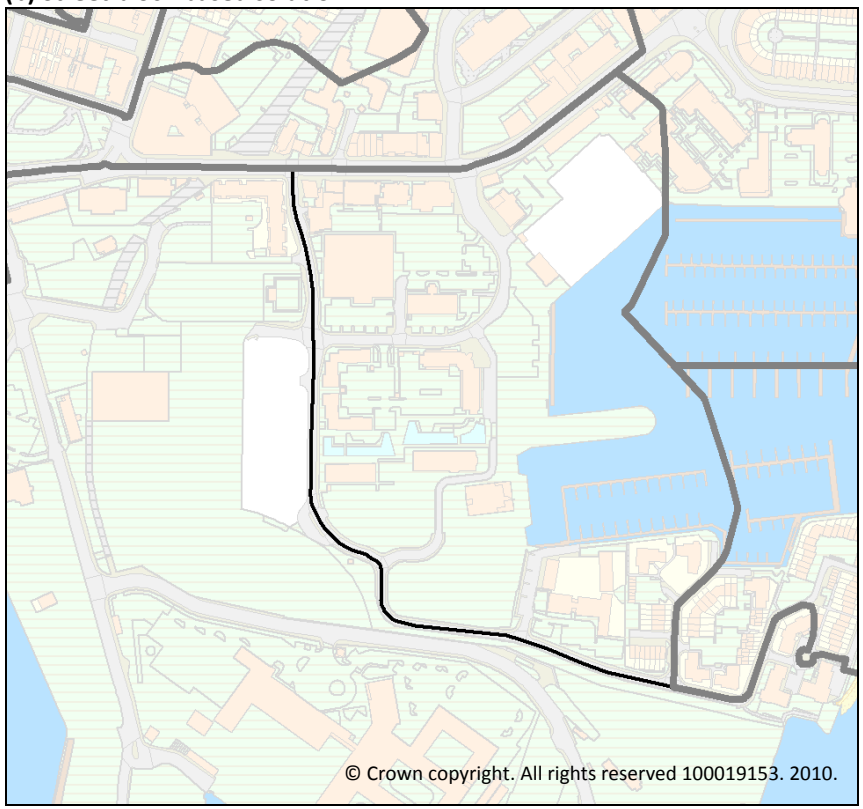
Number of postcodes overlapped with	All	Anglesey	Camden	Lancaster	Liverpool	Manchester	Southampton
1	35.7	29.1	31.0	35.6	40.9	33.7	34.1
2	24.4	21.0	23.9	24.2	26.3	23.4	24.4
3	16.2	16.2	17.7	15.5	15.9	16.0	17.2
4	9.7	10.7	11.6	9.8	7.7	10.8	10.5
5	5.3	7.9	5.7	5.8	4.0	5.9	5.4
6-10	7.8	13.3	8.9	8.1	5.0	8.9	7.9
> 10	0.9	1.8	1.2	1.0	0.3	1.3	0.4

Figure 5 Over-threshold OA in Southampton: comparison of solutions produced by postcode- and street block-based building blocks

(a) Postcode-based solution



(b) Street block-based solution



2001 OA

Maintained OA

4.6.2 Impact on data linkage

Postcodes were used as the building blocks for the 2001 output geographies because they provided small, space-filling, geographical units, built up from addresses, which could be used to create geographies which had an inherent degree of socio-economic homogeneity. They were also a useful linking geography because they enabled users to link their own postcoded data to census data. However, postcodes are continually introduced, terminated and recycled in response to changing household distributions; they also do not have a defined set of boundaries as they are lists of (mail) delivery points. It is important therefore to understand the extent to which the one-to-one match between postcodes and OAs (which was aimed for in 2001) has deteriorated since 2001, as this will influence the accuracy of any postcode to OA look-ups/links published from the 2011 Census data. This will also have an important bearing on the decision as to whether all parts of a postcode should still be required to fall within one maintained 2011 OA (as they were in 2001).

Tables 12 to 14 present the percentage of 2001 postcodes whose constituent dwelling spaces fell within one or more 2001 OAs, LSOAs and MSOAs respectively. Overall, 98.7% of postcodes within the study areas had all of their constituent dwelling spaces falling within one OA, which is not surprising given that the 2001 design process aimed to ensure that all dwelling spaces within a postcode fell within one OA wherever possible. The percentages were notably lower in the rural study areas, presumably due in part to the presence of larger, sprawling, postcodes. In all of the study areas, there was a slightly lower one-to-one match at the OA level than at the LSOA and MSOA levels: this is undoubtedly a scale effect.

Tables 15 to 17 present the same statistics but this time for 2007 postcodes within the 2001 output geographies. Overall, the percentage of postcodes with their constituent dwelling spaces split across more than one 2001 OA had increased from 1.3% in 2001 to 6.3% in 2007. This supports the hypothesis that the (near) one-to-one postcode to OA relationship which existed in 2001 has been degrading through time. This raises three important questions: Is the criterion that, where possible, all parts of any split postcode should be placed within a single OA still a valid design requirement for 2011? What degree of confidence can be placed in any contemporary published one-to-one postcode to OA look-ups? And, on balance, are postcodes the most appropriate set of building blocks for use in maintaining the 2001 output geographies, or do street blocks provide a more useful alternative?

(i) Is the criterion that, where possible, all parts of any split postcode should be placed within a single OA still a valid design requirement?

Tables 18 to 19 present the same statistics as in the previous analyses but this time for the 2007 postcodes within the two sets of maintained postcode- and street block-based output geographies. Unlike the OAPS process, for the purposes of this research, we did not insist that all postcodes should fall within one OA when maintaining the output geographies using our 2007 test data. The percentages of one-to-one matches were very similar both before and after the maintenance process, suggesting that not placing all parts of a split postcode within one maintained OA had very little effect on the ability to link data via a postcode to OA link. Of course, while our study areas were selected because they were representative of the types of population change being experienced across the country, the numbers of split postcodes and maintained areas within them were relatively low, so any effects of split postcodes were always going to be small. Insisting that all parts of a split postcode are placed entirely within one maintained OA imposes significant constraints on the automated design process, thus reducing the number of under- and over-threshold areas for which solutions can be found and leading to less compact shapes (as occurred in the 2001 OAPS process). On balance, therefore, it is suggested that the 2011 automated

maintenance procedures do *not* employ the constraint that all parts of a split postcode must be placed within one maintained OA.

(ii) What degree of confidence can be placed in any published one-to-one postcode to OA look-ups?
In 2001, 98.7% of postcodes matched to one unique OA; by 2007 this percentage had fallen to 93.7% and, presumably, this percentage will have fallen still further by 2011. Nonetheless, these statistics suggest that, in the vast majority of areas, users should still be able to uniquely match postcoded data to one OA. Of course, problems with matching may occur in specific areas if the one-to-many postcode to OA relationships are geographically concentrated.

Tables 20 and 21 present an alternative measure of the degree of certainty associated with matching postcoded data to 2001 OAs, for 2001 and 2007 postcodes respectively. These tables show the absolute percentage differences between the postcode population assigned to each OA via a whole-postcode allocation methodology (whereby the entire postcode's population is redistributed to an OA based on the postcode-OA look-up provided by ONS in 2001 and by the National Statistics Postcode Directory (NSPD) in 2007[3]) compared to a best-fit methodology (whereby a postcode's population is redistributed to OAs based on the individual locations of all of its constituent dwelling spaces). This absolute percentage difference is a measure of the relative accuracy of the two methods.

Table 20 shows that in all study areas combined there was no difference in the populations allocated by the whole-postcode and best-fit methods in approximately 90% of OAs in 2001, and that this percentage was greater than 94% in Camden, Manchester and Southampton. The whole-postcode methodology was least accurate (when compared to the best-fit methodology) in Anglesey and, to a lesser extent, Lancaster, presumably due to the presence of larger rural postcodes which intersect with more OAs than in the urban areas. Even in Anglesey though, over 85% of OAs had less than a 5% difference resulting from the two methods. This close correspondence between the results of the two methods is due to the near one-to-one match between postcodes and OAs in 2001, which in turn resulted from the fact that, where possible, postcodes were required to nest wholly within OAs.

Table 21 shows that by 2007 this close correspondence between the two methods had declined fairly dramatically in all study areas. The mean absolute percentage difference between the two methods across all study areas increased from 0.57 in 2001 to 2.18 in 2007. By 2007, only 52% of OAs across all study areas recorded exactly the same population from the two methods, compared to 90% in 2001. But, overall, approximately 90% of OAs in all study areas still had at least a 95% correspondence between the two methods, indicating that the differences were still small in the majority of areas. All study areas experienced a decrease in the correspondence between the two methods, with Camden seeing the greatest reduction. Notably, even Anglesey, which was included as a relatively stable area, saw a decline in the correspondence. This suggests that the decline is not only due to changes in the intersection of postcodes and OAs due to population change, but possibly also due to reconfigurations of the postcode geography in specific areas. It is also possible that differences in how the postcode to OA look-ups were derived for the whole-postcode allocation methods in 2001 and 2007 (see footnote 1) mean that the two sets of statistics are not directly comparable, although this should not be the case.

[3] Note that the 2001 ONS postcode-OA look-up was produced using population-weighted centroids derived from ONS' own (census) address register; the 2007 NSPD look-up employed a point-in-polygon method based on the postcode centroid and OA boundaries, with the postcode centroid being located within the building of the address which is nearest to the mean location of the addresses within the postcode. These methods are slightly different and will produce slightly different results in some places even if there has been no change to the address locations within a postcode between 2001 and 2007; in the majority of cases though the results should be directly comparable.

These results suggest that the accuracy of any published postcode to OA look-ups which are based on a whole-postcode allocation methodology (such as that currently employed by the NSPD point-in-polygon methodology) had declined between 2001 and 2007 and will presumably have declined still further by 2011.

(iii) On balance, are postcodes the most appropriate set of building blocks for use in maintaining the 2001 output geographies, or do street blocks provide a more useful alternative?

One purported strength of using postcodes rather than street blocks as building blocks for the splitting of over-thresholds OAs, is that it should result in a higher percentage of one-to-one matches between postcodes and OAs. This should aid the linkage of postcoded data with census data. Tables 18 and 19 reveal that, whilst there were very small differences in the percentages of one-to-one postcode to OA matches between the postcode-maintained OAs and the street block-maintained OAs, these differences were nowhere near as large as might have been predicted, although it should be noted that these statistics are based on small numbers of maintained geographies. These results imply that using street blocks as the building blocks, rather than postcodes, would not lead to a significant reduction in a user's ability to link postcoded data with the maintained output geographies, as had been hypothesised. However, despite the fact that the street blocks also produce more compactly-shaped maintained areas, overall postcodes allow more under- and over-threshold areas to be resolved, and generally result in output geographies with superior statistical qualities than those produced using street blocks. It is therefore recommended that ONS employs postcodes as the building blocks for the 2011 automated maintenance procedures.

Table 12 Percentage of 2001 postcodes whose constituent dwelling spaces fall within specified number of 2001 OAs

	Percentage of postcodes within study area						
Number of 2001 OAs which postcode's dwelling spaces fall within	Anglesey	Camden	Lancaster	Liverpool	Manchester	Southampton	Total
1	93.5	99.6	97.2	98.5	99.6	99.8	98.7
2	6.1	0.4	2.8	1.5	0.4	0.2	1.3
3	0.4	0.0	0.0	0.0	0.0	0.0	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 13 Percentage of 2001 postcodes whose constituent dwelling spaces fall within specified number of 2001 LSOAs

	Percentage of postcodes within study area						
Number of 2001 LSOAs which postcode's dwelling spaces fall within	Anglesey	Camden	Lancaster	Liverpool	Manchester	Southampton	Total
1	94.4	99.6	98.0	98.5	99.6	99.8	98.8
2	5.3	0.4	2.0	1.5	0.4	0.2	1.1
3	0.4	0.0	0.0	0.0	0.0	0.0	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 14 Percentage of 2001 postcodes whose constituent dwelling spaces fall within specified number of 2001 MSOAs

	Percentage of postcodes within study area						
Number of 2001 MSOAs which postcode's dwelling spaces fall within	Anglesey	Camden	Lancaster	Liverpool	Manchester	Southampton	Total
1	97.8	99.7	99.2	99.3	99.8	99.8	99.4
2	2.1	0.3	0.8	0.7	0.2	0.2	0.6
3	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 15 Percentage of 2007 postcodes whose constituent dwelling spaces fall within specified number of 2001 OAs

	Percentage of postcodes within study area						
Number of 2001 OAs which postcode's dwelling spaces fall within	Anglesey	Camden	Lancaster	Liverpool	Manchester	Southampton	Total
1	88.9	95.5	93.8	92.1	93.9	96.8	93.7
2	10.1	4.5	6.1	7.4	6.0	3.1	6.0
3	0.9	0.0	0.1	0.5	0.2	0.1	0.3
4	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 16 Percentage of 2007 postcodes whose constituent dwelling spaces fall within specified number of 2001 LSOAs

	Percentage of postcodes within study area						
Number of 2001 OAs which postcode's dwelling spaces fall within	Anglesey	Camden	Lancaster	Liverpool	Manchester	Southampton	Total
1	93.9	97.9	96.7	95.7	97.2	98.2	96.8
2	5.6	2.1	3.2	4.2	2.8	1.8	3.1
3	0.5	0.0	0.0	0.1	0.0	0.0	0.1
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 17 Percentage of 2007 postcodes whose constituent dwelling spaces fall within specified number of 2001 OAs

	Percentage of postcodes within study area						
Number of 2001 OAs which postcode's dwelling spaces fall within	Anglesey	Camden	Lancaster	Liverpool	Manchester	Southampton	Total
1	97.7	99.2	98.9	98.2	98.8	99.3	98.7
2	2.2	0.8	1.1	1.8	1.2	0.7	1.3
3	0.1	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 18 Percentage of 2007 postcodes whose constituent dwelling spaces fall within specified number of maintained (postcode-based) OAs

	Percentage of postcodes within study area						
Number of maintained (postcode) OAs which postcode's dwelling spaces fall within	Anglesey	Camden	Lancaster	Liverpool	Manchester	Southampton	Total
1	88.9	95.5	93.8	92.0	93.6	96.8	93.6
2	10.1	4.5	6.1	7.6	6.1	3.2	6.1
3	0.9	0.0	0.1	0.5	0.2	0.1	0.3
4	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 19 Percentage of 2007 postcodes whose constituent dwelling spaces fall within specified number of maintained (street block-based) OAs

	Percentage of postcodes within study area						
Number of maintained (street block) OAs which postcode's dwelling spaces fall within	Anglesey	Camden	Lancaster	Liverpool	Manchester	Southampton	Total
1	88.7	95.5	93.8	91.9	93.5	96.8	93.6
2	10.2	4.5	6.1	7.6	6.2	3.1	6.1
3	0.9	0.0	0.1	0.5	0.3	0.1	0.3
4	0.2	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 20 Distribution of percentage differences between postcode population assigned to OA by whole postcode assignment* and best-fit^ methodologies, for 2001 postcodes

Absolute % difference from best-fit population	Percentage of OAs within study area						
	Anglesey	Camden	Lancaster	Liverpool	Manchester	Southampton	All
>20	1.3	0.0	0.7	1.0	0.1	0.0	0.5
10 to 20	3.5	0.1	1.3	2.1	1.0	0.0	1.2
5 to 10	7.5	0.4	5.3	2.3	1.6	0.8	2.1
0 to 5	44.9	4.8	14.3	5.4	2.2	2.2	6.6
0	42.7	94.7	78.5	89.2	94.9	97.0	89.6
Mean absolute % difference	2.36	0.13	0.97	0.85	0.33	0.10	0.57

* Whole postcode assignment: entire postcode's population is redistributed to the OA specified in the 2001 ONS postcode-OA look-up table

^ Best-fit method: postcode's population redistributed to OAs based on locations of its constituent dwelling spaces

Table 21 Distribution of percentage differences between postcode population assigned to OA by whole postcode assignment* and best-fit^ methodologies, for 2007 postcodes

Absolute % difference from best-fit population	Percentage of OAs within study area						
	Anglesey	Camden	Lancaster	Liverpool	Manchester	Southampton	All
>20	2.2	2.6	0.9	1.5	1.1	1.5	1.5
10 to 20	6.6	1.9	3.5	5.5	3.8	0.8	3.7
5 to 10	15.9	5.5	6.8	7.8	7.2	3.4	6.9
0 to 5	57.7	32.7	40.4	37.3	37.7	26.4	36.4
0	17.6	57.3	48.4	47.9	50.2	67.8	51.5
Mean absolute % difference	3.80	2.37	1.85	2.47	1.94	1.53	2.18

* Whole postcode assignment: entire postcode's population is redistributed to the OA specified in the 2007 NSPD

^ Best-fit method: postcode's population redistributed to OAs based on locations of its constituent dwelling spaces

5 Limitations and other points to note

Population estimates: the mid-year population estimates employed were the most up-to-date available and were provided for the research team by ONS at the postcode level, which is at a lower resolution than is usually available for public/academic use. Despite this, they are still only estimates: the magnitude and spatial distribution of any patterns observed may not be accurate. In addition, it is impossible to predict whether the trends observed will continue or change between 2007/8 and 2011.

Concurrency of datasets: whilst every effort has been made to employ datasets which are as concurrent with one another as possible, there are inevitably slight differences in the dates that the various datasets represent. Table 22 summarises the various datasets employed in this research.

Table 22 Datasets employed in this research

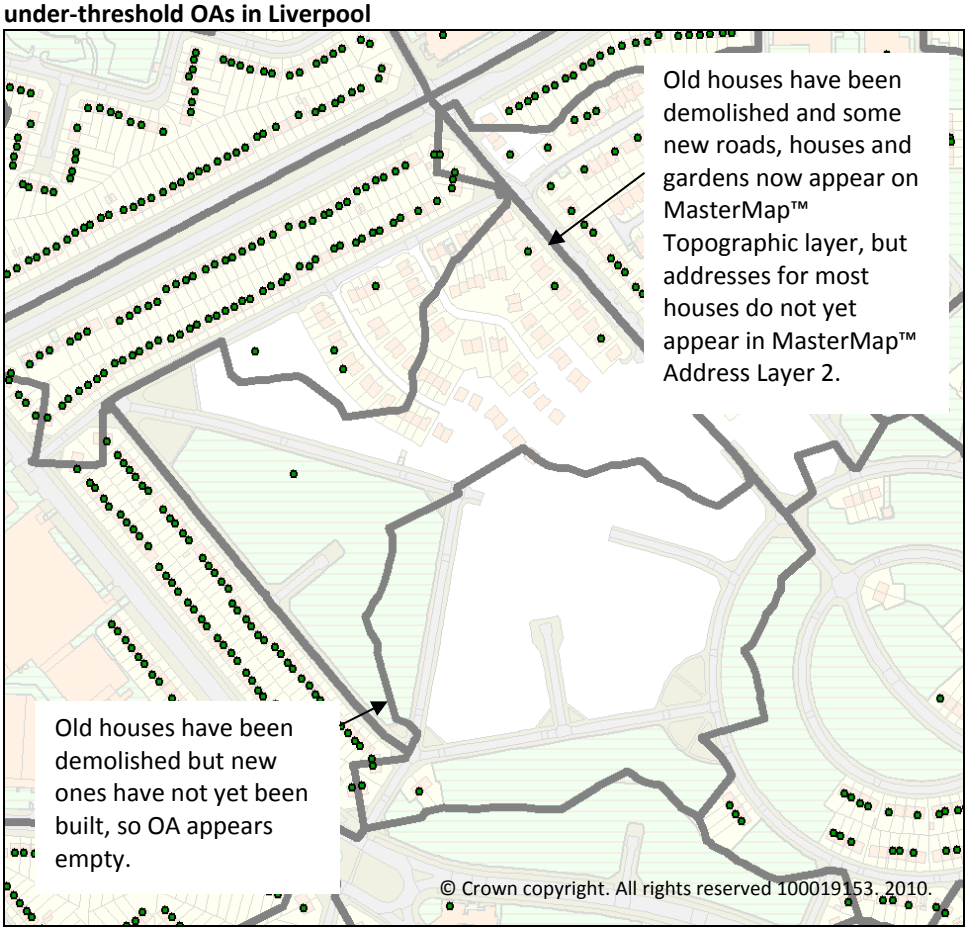
Dataset	*Date*	*Source*
MasterMap Address Layer 2	March 2008	ONS (Ordnance Survey product)
2001 Census FORMID data	2001	ONS
2001 Output Area, Lower Layer Super Output Area, Middle Layer Super Output Area boundaries	2001	UKBORDERS
Mid-year population estimates for postcodes	Mid 2007	ONS
Special population counts and look-up tables	Mid 2007	ONS
MasterMap™ Integrated Transport Network and Topo	December 2007	UKBORDERS (Ordnance Survey product)
MasterMap™ Topographic Layer	December 2007	UKBORDERS (Ordnance Survey product)
Ordnance Survey Meridian 2 (railways)	October 2008	UKBORDERS
National Statistics Postcode Directory	February 2008	UKBORDERS
Universities UK Student Residences list	March 2009	www.universitiesuk.ac.uk

These differences in the currency of datasets will lead to inevitable differences in the completeness and accuracy of the locations and attributes of features such as addresses, roads and postcodes. This should be borne in mind when interpreting apparent changes in the size and spatial location of populations and households.

Changes occurring on Census night: an inevitable problem with a cross-sectional survey such as the Census, is that changes which are still in process on Census night can distort the patterns observed. Such changes are often further conflated by differences in the concurrency of datasets. An example of both the issues of on-going change and discrepancies in the concurrency of datasets can be seen in a group of under-threshold OAs in Liverpool. Figure 6 shows two OAs which were within-threshold in 2001 but have since seen the subsequent vacating of premises and demolition of old housing and construction of new housing and associated infrastructure. Examination of the 2001 Census data, 2007 test data, and the roads and buildings in the area, reveals that the various datasets show this redevelopment at different stages of completion, with houses having been built, address points and postcodes having been assigned to some but not all of the houses, and not all roads yet being shown. Here, the difficult task of capturing the process of change is further conflated by time lags in the various datasets mapping this change. The Census, as a cross-sectional survey, simply takes a snap-shot of the country and its population at that time and therefore, quite rightly, does not attempt to reconcile or predict such changes. But, because there is an evitable time lag in features appearing on the various datasets employed in the design of the output geographies, areas can be designed or maintained on the basis of incomplete or inaccurate information, as will have occurred in this case, where the currently under-threshold OAs were

correctly merged to make them within-threshold, but which would soon see population growth when people moved into the new properties.

Figure 6 Example of issues raised by on-going change and discrepancies in the concurrency of datasets: under-threshold OAs in Liverpool



Method used to allocate population, household count, tenure and accommodation type onto dwellings: the method employed to allocate populations, households, tenure and accommodation type onto dwellings within the study areas was detailed in previous reports to OAMPAG. There are a number of assumptions built into this methodology, which may have resulted in inaccurate allocations. In particular, the allocation of population to CEs, which was the final stage of the allocations, has the highest degree of uncertainty associated with it.

Uncertainty surrounding the recording of CEs: It is difficult to directly compare changes in the number of CEs between 2001 and 2007/08 due to inconsistencies between datasets in the way that some CE types, such as student residences, were recorded. For example, in some cases, student residences in the 2007/8 AddressLayer 2 dataset were classified as multiple occupancy CEs all with the same grid reference, whereas in the 2001 Census counts the same residences were recorded as one CE on that grid reference. This has the effect of artificially inflating the count of the number of CEs within OAs, even if the CE and number of residents has not changed between 2001 and 2007/08 (as in the maximum CE counts of 629 and 415 in Camden and Liverpool respectively). In other

instances (e.g. for the maximum CE count of 245 in Manchester), there was a change (between 2001 and 2007/08) in whether the dwelling space was recorded as a CE or as a residential property.

Urban/rural constraint: We did not use an urban/rural constraint when aggregating/merging areas because there was no urban/rural dataset readily available for 2007/8. ONS needs to decide whether to use such a constraint and, if so, what measure of urban/rural to employ.

Samantha Cockings and Andrew Harfoot (September 2010)

APPENDICES

APPENDIX A

Tables A1(i-vi): Threshold breaches in 2007, classified by type of breach and reason for over-threshold breaches, by output geography level, by study area

A1(i) Anglesey

	Total	Under threshold	Within threshold	Over threshold			
				As at 2001			Other reasons
					Resid pop below upper threshold		
OAs	227	1	225	0	0	0	1
LSOAs	44	0	44	0	0	0	0
MSOAs	9	0	9	0	0	0	0

A1(ii) Camden

	Total	Under threshold	Within threshold	Over threshold			
				As at 2001			Other reasons
					Resid pop below upper threshold		
OAs	734	4	715	0	5	6	4
LSOAs	133	1	132	0	0	0	0
MSOAs	28	1	27	0	0	0	0

A1(iii) Lancaster

	Total	Under threshold	Within threshold	Over threshold			
				As at 2001			Other reasons
					Resid pop below upper threshold		
OAs	455	0	452	0	2	0	1
LSOAs	89	1	87	0	1	0	0
MSOAs	19	0	19	0	0	0	0

A1(iv) Liverpool

	Total	Under threshold	Within threshold	Over threshold			
				As at 2001			Other reasons
					Resid pop below upper threshold		
OAs	1501	30	1445	1	5	2	19
LSOAs	291	6	284	0	0	0	1
MSOAs	59	0	59	0	0	0	0

A1(v) Manchester

	Total	Under threshold	Within threshold	Over threshold			
				As at 2001			Other reasons
					Resid pop below upper threshold		
OAs	1341	8	1282	1	3	1	46
LSOAs	259	0	251	0	1	1	6
MSOAs	53	0	52	0	0	0	1

A1(vi) Southampton

	Total	Under threshold	Within threshold	Over threshold			
				As at 2001			Other reasons
					Resid pop below upper threshold		
OAs	730	0	717	0	3	1	9
LSOAs	146	4	140	0	0	0	2
MSOAs	32	0	32	0	0	0	0

APPENDIX B

Table B1 Resolved and unresolved 2001 OAs (using postcode-based building blocks)

LA Code	Count	Over-threshold	Resolved	Unresolved				Under-threshold	Resolved	Unresolved	
				Rule 1	Rule 2	Rule 3	Other			No BB	Other
Anglesey	227	1	1	0	0	0	0	1	1	0	0
Camden	734	15	10	0	4	0	1	4	4	0	0
Lancaster	455	3	1	0	1	0	1	0	0	0	0
Liverpool	1501	26	18	1	4	0	3	30	30	0	0
Manchester	1341	51	44	1	6	0	0	8	8	0	0
Southampton	730	13	10	0	2	0	1	0	0	0	0
All	4988	109	84	2	17	0	6	43	43	0	0

Table B2 Resolved and unresolved 2001 OAs (using street block-based building blocks)

LA Code	Count	Over-threshold	Resolved	Unresolved				Under-threshold	Resolved	Unresolved	
				Rule 1	Rule 2	Rule 3	Other			No BB	Other
Anglesey	227	1	1	0	0	0	0	1	1	0	0
Camden	734	15	6	0	7	2	0	4	4	0	0
Lancaster	455	3	0	0	3	0	0	0	0	0	0
Liverpool	1501	26	13	1	12	0	0	30	30	0	0
Manchester	1341	51	35	1	12	1	2	8	8	0	0
Southampton	730	13	4	0	7	1	1	0	0	0	0
All	4988	109	59	2	41	4	3	43	43	0	0

Table B3 Resolved and unresolved 2001 LSOAs

LA Code	Count	Over-threshold	Resolved	Unresolved				Under-threshold	Resolved	Unresolved	
				Rule 1	Rule 2	Rule 3	Other			No BB	Other
Anglesey	44	0	0	0	0	0	0	0	0	0	0
Camden	133	0	0	0	0	0	0	1	0	0	1
Lancaster	89	1	0	1	0	0	0	1	0	0	1
Liverpool	291	1	0	0	0	0	1	6	6	0	0
Manchester	259	8	5	2	0	0	1	0	0	0	0
Southampton	146	2	1	0	0	0	1	4	4	0	0
All	962	12	6	3	0	0	3	12	10	0	2

Table B4 Resolved and unresolved 2001 MSOAs

LA Code	Count	Over-threshold	Resolved	Unresolved				Under-threshold	Resolved	Unresolved	
				Rule 1	Rule 2	Rule 3	Other			No BB	Other
Anglesey	9	0	0	0	0	0	0	0	0	0	0
Camden	28	0	0	0	0	0	0	1	1	0	0
Lancaster	19	0	0	0	0	0	0	0	0	0	0
Liverpool	59	0	0	0	0	0	0	0	0	0	0
Manchester	53	1	1	0	0	0	0	0	0	0	0
Southampton	32	0	0	0	0	0	0	0	0	0	0
All	200	1	1	0	0	0	0	1	1	0	0

APPENDIX C

Tables C1(i-vi) Statistical characteristics of (bottom-up) maintained OAs (using postcodes and street blocks), compared to 2001 and 2007 data in 2001 OAs, by study area

C1(i) Anglesey

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	227	294.4	53.7	124.9	17.4	0.114	0.117	44.22	16.27
2007	227	304.0	79.9	128.0	32.4	0.103	0.118	44.22	16.27
Bottom-up maintained (postcodes)	228	302.6	67.0	127.4	22.0	0.104	0.121	44.17	16.07
Bottom-up maintained (street blocks)	228	302.6	67.3	127.4	22.2	0.104	0.120	44.04	16.09

C1(ii) Camden

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	734	269.8	74.2	124.8	14.8	0.130	0.148	34.21	12.22
2007	734	316.0	113.6	124.6	33.6	0.100	0.140	34.21	12.22
Bottom-up maintained (postcodes)	738	314.2	104.9	123.9	32.4	0.101	0.137	34.23	12.30
Bottom-up maintained (street blocks)	734	316.0	109.1	124.6	32.7	0.101	0.137	34.23	12.25

C1(iii) Lancaster

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	455	294.3	190.1	122.7	14.2	0.179	0.259	39.05	14.24
2007	455	315.4	211.4	121.6	23.0	0.155	0.236	39.05	14.24
Bottom-up maintained (postcodes)	456	314.8	210.5	121.4	21.8	0.155	0.236	39.02	14.23
Bottom-up maintained (street blocks)	455	315.4	211.4	121.6	23.0	0.155	0.236	39.05	14.24

C1(iv) Liverpool

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	1501	292.8	82.9	125.2	18.5	0.161	0.245	38.06	15.31
2007	1501	290.1	112.1	122.4	43.0	0.142	0.231	38.06	15.31
Bottom-up maintained (postcodes)	1495	291.3	103.3	122.9	33.2	0.142	0.232	38.10	15.41
Bottom-up maintained (street blocks)	1485	293.3	103.3	123.7	35.2	0.142	0.231	38.00	15.23

C1(v) Manchester

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	1341	292.9	103.3	124.9	16.5	0.184	0.205	38.26	14.17
2007	1341	341.6	160.1	138.2	57.6	0.152	0.168	38.26	14.17
Bottom-up maintained (postcodes)	1410	324.9	140.4	131.4	29.2	0.154	0.173	38.11	14.22
Bottom-up maintained (street blocks)	1384	331.0	147.4	133.9	45.7	0.153	0.172	37.93	14.06

C1(vi) Southampton

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	730	297.9	88.2	125.0	12.9	0.184	0.194	37.44	13.73
2007	730	316.7	130.3	125.9	35.5	0.167	0.187	37.44	13.73
Bottom-up maintained (postcodes)	747	309.5	115.2	123.0	22.8	0.168	0.188	37.34	13.66
Bottom-up maintained (street blocks)	735	314.6	128.4	125.1	33.4	0.167	0.187	37.32	13.71

Tables C2(i-vi) Statistical characteristics of (bottom-up) maintained LSOAs, compared to 2001 and 2007 data in 2001 LSOAs, by study area

C2(i) Anglesey

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	44	1518.8	341.5	644.5	141.7	0.042	0.048	43.39	15.95
2007	44	1568.3	369.6	660.4	157.8	0.038	0.050	43.39	15.95
Bottom-up maintained	44	1568.3	369.6	660.4	157.8	0.038	0.050	43.39	15.95

C2(ii) Camden

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	133	1488.9	86.5	688.7	103.9	0.086	0.079	42.29	13.37
2007	133	1743.7	198.7	687.5	119.3	0.062	0.070	42.29	13.37
Bottom-up maintained	133	1743.7	198.7	687.5	119.3	0.062	0.070	42.29	13.37

C2(iii) Lancaster

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	89	1504.7	443.8	627.4	112.9	0.117	0.143	43.00	15.23
2007	89	1612.7	486.6	621.8	118.1	0.099	0.128	43.00	15.23
Bottom-up maintained	89	1612.7	486.6	621.8	118.1	0.099	0.128	43.00	15.23

C2(iv) Liverpool

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	291	1510.2	127.6	645.6	93.1	0.118	0.149	41.51	15.18
2007	291	1496.5	229.5	631.5	124.2	0.104	0.146	41.51	15.18
Bottom-up maintained	286	1522.7	235.6	642.5	124.7	0.104	0.145	41.35	14.83

C2(v) Manchester

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	259	1516.7	182.2	646.5	95.5	0.128	0.106	43.45	15.53
2007	259	1768.9	457.9	715.5	245.2	0.104	0.099	43.45	15.53
Bottom-up maintained	266	1722.3	331.6	696.6	136.9	0.105	0.101	43.62	15.42

C2(vi) Southampton

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	146	1489.3	117.8	624.8	92.8	0.126	0.106	43.71	14.61
2007	146	1583.5	289.0	629.6	135.0	0.112	0.105	43.71	14.61
Bottom-up maintained	143	1616.7	314.2	642.8	130.7	0.111	0.104	43.91	15.06

Tables C3(i-vi) Statistical characteristics of (bottom-up) maintained MSOAs, compared to 2001 and 2007 data in 2001 MSOAs, by study area

C3(i) Anglesey

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	9	7425.4	1234.5	3150.7	594.2	0.018	0.026	68.67	29.39
2007	9	7667.0	1261.5	3228.7	555.3	0.015	0.027	68.67	29.39
Bottom-up maintained	9	7667.0	1261.5	3228.7	555.3	0.015	0.027	68.67	29.39

C3(ii) Camden

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	28	7072.1	783.2	3271.5	436.5	0.055	0.041	42.54	13.76
2007	28	8282.5	1044.9	3265.7	468.2	0.036	0.034	42.54	13.76
Bottom-up maintained	27	8589.2	1583.2	3386.7	501.9	0.036	0.034	43.62	14.30

C3(iii) Lancaster

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	19	7048.1	1768.5	2938.9	737.8	0.061	0.087	44.20	14.44
2007	19	7554.1	1991.8	2912.5	801.1	0.047	0.078	44.20	14.44
Bottom-up maintained	19	7554.1	1991.8	2912.5	801.1	0.047	0.078	44.20	14.44

C3(iv) Liverpool

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	59	7448.7	999.9	3184.2	416.7	0.087	0.092	46.20	14.01
2007	59	7381.2	1210.2	3114.7	540.6	0.076	0.095	46.20	14.01
Bottom-up maintained	59	7381.2	1210.2	3114.7	540.6	0.076	0.095	46.20	14.01

C3(v) Manchester

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	53	7411.7	1028.8	3159.5	382.7	0.078	0.056	44.08	11.88
2007	53	8644.1	1469.0	3496.3	624.2	0.068	0.060	44.08	11.88
Bottom-up maintained	54	8484.0	1501.0	3431.5	520.5	0.068	0.054	44.16	12.05

C3(vi) Southampton

	Count	Mean total population	SD total population	Mean total households	SD total households	IAC tenure	IAC accommodation type	Mean P^2/A	SD P^2/A
2001	32	6795.2	687.9	2850.5	348.1	0.074	0.052	36.65	8.49
2007	32	7224.8	1147.3	2872.4	445.4	0.068	0.055	36.65	8.49
Bottom-up maintained	32	7224.8	1147.3	2872.4	445.4	0.068	0.055	36.65	8.49