

**Census2011Geog:
Executive Summary and Recommendations**

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>



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/categories¹. Suggestions: [Tenure](#): owns; shared; rents; rfree. [Accom type](#): detached; semi; terraced; flat; parhouse; commerce; nonperm.
 - iv) Weighting for each variable? [100](#)
 - d) Shape constraint
 - i) Use or not? [Use](#)
 - ii) What score/measure? [Compactness² = Perimeter²/Area](#)
 - iii) Weighting? [100](#)
 - e) Minimum boundary length
 - i) Use or not? [Use](#)
 - ii) What value? [10%](#)

¹ 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

<i>Suggested category</i>	<i>2011 Code</i>	<i>2011 Description</i>
Detached	1	Detached
Semi	2	Semi-detached
Terraced	3	Terraced (including end-terrace)
Flat	4	In a purpose-built block of flats or tenement
Parhouse	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

<i>Suggested category</i>	<i>2011 Code</i>	<i>2011 Description</i>
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

² AZTool and the research reported here employ perimeter²/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.