

Work Session on Geographical Information Systems
(Brighton, United Kingdom, 22-25 September 1997)

Item (iv) of the provisional agenda

**USE OF GIS FOR PLANNING THE 2001 CENSUS:
ONS PLANS FOR INPUT AND OUTPUT GEOGRAPHIES FOR
ENGLAND AND WALES**

by

Submitted by the Office for National Statistics, United Kingdom¹

¹ Prepared by Chris Denham.

I. INPUT GEOGRAPHY : THE PROPOSED COLLECTION AREA PLANNING SYSTEM

1. The development of a system to plan collection areas which is quite radically different from that used until 1991 has now been completed for the 1997 Census test. It promises to be more cost effective than the previous system, and, provided no unexpected problems are reported from the field in the Census test and provided that there are no severe increases in the cost of 'brought in' components, the test system seems likely to be adopted for the 2001 Census. At this stage, however, the 1991 Census system remains a viable fall-back. A final decision will be made around the end of 1997. In summary, the new system:

- a) uses data from the national mapping agency for Great Britain - Ordnance Survey - in the form of Address Point (which provides a unique co-ordinate reference for every address) and digital map data through a GIS;
- b) produces output for use in the field in the form of a single simplified reference map sheet for each collection area (enumeration district) and an enumerator's record book with pre-listed addresses;
- c) could produce digital enumeration district boundaries if these are required for output geography;

2. The input to the system is 1991 Census enumeration districts (EDs) (in the form of digital boundaries) plus the Ordnance Survey products - Address-Point, Boundary-Line (current statutory boundaries held in digital form), and 1:10,000 scale raster maps - to enable EDs to be created. The system uses a network of computers running customised GIS software, and is known as GAPS (Geographical Area Planning System).

3. In the system a 1991 Census ED identity is automatically added to each address on the Address-Point database (by a simple computer process of identifying all Address-Point 1 metre grid references falling within the 1991 ED polygons). This allows revised counts of addresses for 1991 EDs to be established, and, by using look-up tables for ED grading information, a provisional set of 2001 EDs is produced. The system then highlights EDs that require amendment to produce an enumeration workload between determined limits, and allows revised boundaries to be created on screen.

4. Any redesigning of EDs necessary to meet planning criteria is accomplished by the identification of 'blocks' of addresses to be moved between EDs. This is achieved by digitising, on screen, a temporary polygon and interrogating the Address-Point database to find the number of addresses in the polygon.

5. The system removes the need to create extensive folios of paper maps and to overlay statutory boundaries - raster maps and Boundary-Line automatically achieve this. Only changed EDs will new boundaries, and these can be drawn interactively on screen more quickly than drawing by hand. An automatic by-product would be digital boundaries for 2001 Census EDs. Map 'collation' work is carried out within the system, and customised maps are produced for enumerators as single sheets (at A4 or A3 size at various scales) via colour laser printers. Composite maps for use by the field managers are produced by larger colour plotters.

6. Enumerators in the 1997 Test are also being provided with lists of addresses within their EDs to supplement the reference maps. These are pre-printed in the enumerators' record books. In addition, the new system would create the computer files of geographical information necessary to process the census forms and produce the output.

7. The performance of the new system will be monitored throughout the 1997 Census test, not only the technical aspects of GAPS but also whether or not, for example, the simplified reference maps or pre-listing of addresses have impacts on completeness of coverage. One interesting geographical aspect is that there are insufficient data to organise the pre-listed addresses into spatially optimum walks, and it will be instructive to see how far enumerators follow the listed order. Adoption of the system for 2001, and any further modifications, will depend on performance, but may also depend on the costs of the 'brought-in' data for the whole of England and Wales to realise the gain in cost-effectiveness demonstrated in the trial stages.

II. OUTPUT GEOGRAPHY : OPTIONS

8. Output geography depends on the basic spatial unit to which data are referenced both to underpin Census output for statutory and geo-statistical areas and to provide the basis for local and small area statistics. The current options in England and Wales are:

- a) enumeration districts (the Census collection areas);
- b) postcode units (the smallest unit, numbering some 1.6 million in the UK, in the hierarchical system of postal geography used principally for the sorting and delivery of mail);
- c) unique grid references of addresses (provided by Address Point) ; or
- d) a combination of any of these.

9. Options (ii) or (iii) would give the separation of input and output geography and scope for considerable flexibility. However, unless all users from all customer sectors agree on either option (i) or (ii) (and on a single aggregating of the basic units within the postcode option), and certainly in the case of option (iii), the issue arises of dealing with alternative geographies and any consequent risk of disclosure of information about identifiable individuals contrary to any assurances given or any statutory obligations. Whatever options are developed, there are also likely to be directories or 'look-up tables' linking various current and historic geographies.

III. ALTERNATIVE GEOGRAPHIES

10. A logically strong case has been put forward about the disclosure risks from production of output for any combination of small area building bricks. However, statistical data can be modified and the number of area combinations can be limited. This was the situation for the 1991 Census in England and Wales where modified statistics for EDs and postcode sectors were supplied, giving two overlapping boundary sets but no further user choice. No problem with apparent disclosure though differencing has been reported.

11. Where small area statistics are supplied for alternative geographies, as they were for England and Wales from the 1971, 1981 and 1991 Censuses, but where the user has no further choice, a process of disclosure through iterative differencing is not possible in any systematic way. Further protection against risks of disclosure could be given, for example, by modifying records on the output bases (provided that this is acceptable under Data Protection legislation) or by using 'non-integer' forms of statistics. At this stage ONS does not see the need to rule out the option of alternative geographies if they are needed for 2001, and is prepared to consider the trade-offs between geographical flexibility and statistical detail to meet user requirements.

IV. POSTCODE BUILDING BRICKS

12. The separation of output geography from collection geography in England and Wales through building output areas from unit postcodes is seen as a leading option at the moment. However, ONS does not have the resources to replicate the process established in Scotland with

manually mapped postcode boundaries, and ‘output areas’ produced would not necessarily follow the form and size of those produced in Scotland for the 1991 Census. A possible method in England and Wales, now successfully trialled as a prototype, would be:

- a) to use Ordnance Survey Address-Point data, particularly if they are obtained for ED planning and pre-listing for enumeration, to generate synthetic polygons around each ‘segment’ of co-ordinate references representing the addresses in the postcodes;
- b) after enumeration, to fit postcodes and addresses enumerated but not pre-listed into the geographic base, modifying the postcode unit polygons as necessary;
- c) if population thresholds are critical, to prepare final populations for each polygon/building brick after enumeration; and
- d) to group the polygon building bricks automatically to output areas within any boundary, contiguity, or other constraints which it is possible to introduce into the system to meet user requirements and to make boundaries as ‘natural’ as possible.

13. Particular boundary constraints could be to wards, civil parishes, parliamentary constituencies, National Parks and so on, or to groups of EDs from previous censuses to give area continuity. This would involve the splitting of postcode units, and ONS would require cases to be made for requirements for exact fit rather than best fit by postcode units.

14. The order of the stages (I) to (iv) could be interchanged to some extent, and the final grouping could be to a single set of output areas, to a hierarchy for different levels of output, to alternative geographies, or to any groupings specified by customers. Some geographical inaccuracy would enter at stage (ii) unless each ‘new’ address was given a co-ordinate reference and any provisional polygons were revised accordingly.

15. Initial trials of methods have been successful, and the next stage is to establish feasibility for what could be an activity carried out at a critical stage of output production, and also to establish costs (all or part of which may have to be passed on to customers). Issues which have to be taken into consideration are:

- a) the additional costs of the initial use of Address-Point information which have to be negotiated with Ordnance Survey, together with consideration of any royalty for ongoing use with the statistical data;
- b) the additional costs of the output area system, and any costs of ‘customisation’;
- c) the acceptability of synthetic boundaries;
- d) the fact that customer involvement in the definition of a single set of output areas would be limited by the methods necessary to cope with zoning over one million basic building bricks in a limited time period;
- e) the error introduced by, or the costs of referencing, ‘new’ addresses; and
- f) the maintenance of the changing postcode set against the ‘frozen’ 2001 Census boundaries, and, indeed, the question of whether ONS should simply ‘buy in’ a set of boundaries generated by some other agency, accepting the loss of any ‘customisation’ for census purposes and the inevitable mis-match with addresses as enumerated in the Census.

16. Arrangements have been made with Ordnance Survey to address the issues arising from use of their data, and ONS is also discussing the acceptability of such output geography with customers.

17. In summary, ONS will have to determine whether the vast bulk of customers would opt for such output areas, particularly if there is constraint to a single set of areas, in preference to EDs.

V. UNIQUE GEOGRAPHICAL REFERENCES

18. The possible use of 1m references taken from Address-Point as the basic spatial unit for output geography is regarded by ONS as a secondary option at the moment. Nevertheless, it would solve many persistent geographical problems, and would provide the scope to build to any other area sets, past, present or future, particularly if adopted as standard for other data. Some impetus is given to the option by the possibility that the emphasis in Royal Mail use of postcodes will shift to unique address codes for delivery sequencing purposes, and the postcode units in their current geographical form may gradually become redundant.

19. With unique referencing, the scope for customer choice of output areas could be infinite and would be in line with the stated objective of increased customisation of output from the 2001 Census, but, whilst disclosure risks are not inherent in the use of co-ordinate references as such, major issues of perceived and actual risks of disclosure would have to be tackled. ONS will be undertaking some initial trials in 1997 of feasibility and costs, particularly to see if there would be major differences from the use of postcodes alone as a basic spatial unit, with a possible scenario of holding postcodes and references on the output data bases to give the potential for short and long term flexibility.