

Outputs, Access and Licensing

Barry Leventhal, Chair, MRS Census & Geodemographics Group

Representing the Association of Census Distributors and the MRS CGG

1. Introduction

This paper has been compiled from views submitted by members of both the ACD and the MRS CGG. Thus it comprises a cross-section of views from major private sector users of Census and related data. ACD members compile, process and add value to Census data, which they then re-supply to their clients; CGG members are a mixture of different types of user, predominantly from the market research industry, but also including users from retail, consultancies, academia, and Census distributors. Both the intermediaries represented within both groups, and the end-users from the CGG, tend to be very experienced in their use of Census data, including many whose first experience was with the 1981 Census – some even earlier.

Thus the experience of Census data usage is extensive. Both the ACD (formerly ACA) and the CGG have had close contact with ONS Census Division (and its predecessor, OPCS) since their inception. The CGG was established in 1989, the ACA in 1993/4. Both organisations have been represented on the Business Advisory Group since it started. Members of the ACD and CGG do not merely have long experience of the use of Census data – this experience is also deep. Among their number are practitioners with vast experience of processing the data, modelling the data, and constructing neighbourhood classifications. They know how it should and should not be used. They have been involved in the consultative process, both within their individual organisations (ACD, CGG and DUG – the Demographic User Group, which is also represented within the CGG) and through BAG, now BAPIAG.

2. Purpose

The purpose in preparing this paper is to record what, in the view of these groups, went well with the processing and publication of the 2001 Census – and what went less well. We also want to look forward to the 2011 Census - the planning of the 2011 Census has already started, but we are concerned that the ONS team in place at the time of taking that Census and its subsequent processing will, inevitably, have changed from the team in place for the 2001 Census. Consequently there is a danger that lessons learned in the period 2001 to date will not be available to the 2011 team. Although this note is mainly addressed to ONS, we believe it is equally applicable to the other Census Offices, GROS and NISRA, for the same reason.

We are aware that ONS has itself reviewed and commented on the performance of the 2001 Census. However, we believe that a view from the user community should also be of value – indeed, in terms of the purpose of the Census itself, should perhaps be more

important! We agree with ONS' own statement that 'investment in collecting statistical information is only worthwhile if the output meets user requirements and provides value' (source: Paper AG(04)08). While the ACD, CGG and DUG are all representative of private sector interests (broadly speaking – although the CGG has wider membership) it is our perception, from discussions with other user sectors, that the views of users across all sectors are very similar.

3. Main issues and views

Our review of the main issues, and our views on them, are summarised in this section and expanded subsequently.

The background to our review, and therefore its context, is the original aims of ONS expressed at an early stage of the planning process for the 2001 Census.

ONS said they intended to deliver 'a quality Census';

- on a timetable that would be published, then adhered to;
- the 'one number Census' would add to the timescale for publication, but would resolve issues of undercounting.

We would agree that, taken as a whole, the 2001 Census was a quality product. However, there were serious delays in the delivery of small-area data (well beyond the original timetable); and it is clear that the one-number Census, although a good idea in itself, did not cope with undercounting in those Metropolitan areas where the problem was most severe.

If these comments seem unduly negative, that is not our intention; we wish to record problems experienced by users of the data, so they may help with future planning. We appreciate the context – far more data were distributed than in previous Censuses, most of those data being correct, all of them being valuable and useful. However, for the organisations and individuals charged with processing the data – especially the OA-level data – there were numerous frustrations, and much wasted resource.

We felt that the planning and consultative process for the 2001 Census was first-class, and a model of its type. While similar in content to the 1991 Census planning process, the 2001 process was more thorough and more inclusive of user groups. All who were involved in the process from the ACD and CGG were pleased by the opportunity to make their input to the output planning process. Proposed formats for tables were circulated and discussed; suggestions were taken on board and implemented if a good case was made. However, late in the process, lack of prior consultation on some issues – notably disclosure control – caused many problems.

4. Outputs, Access & Licensing

Census Access was a major change for the commercial sector which came in train with

the 2001 Census. Previously, while the central and local government, health and academic sectors had central agreements with the Census Offices for 'free at point of use' access to Census data, the commercial sector had 'enjoyed' a royalty regime – which could be very expensive. For example, a commercial organisation that wished to acquire a national ED-level set of (say) 100 1991 Census variables could expect to pay nearly £100,000 in royalties. Census Access changed this situation dramatically, and necessitated a complete revision of the licensing agreement between ONS and the Census VARs.

While all sectors other than the commercial sector tend to have centralised arrangements for Census data dissemination, in practice the Census VARs account for the lion's share of dissemination into the commercial sector. Some end-users would obtain the data directly from the Census Offices, of course, either from Customer Services or via the NeSS website – but still the Census VARs were expecting to serve the bulk of the market. Of course, simple resupply (given free data) would not be very attractive – but the Census distributors add value to the data in various ways (for example, extracting specific sets of variables, supplying them in particular formats, perhaps also supplying bundled software, building neighbourhood classifications – etc.).

Discussions about a new licence agreement to cover the 2001 Census started in 1997 (following the negotiation of the mid-term agreement in 1996). In the event, the adoption of Census Access in 2002 meant that the ACD needed to start again. We agreed with ONS that the click/use licence was not fully adequate to cover the special issues involved in re-supplying Census data; and we embarked upon the discussion of the Distribution Licence. A distribution licence allows the licensee to sub-licence its clients to re-use and publish the Census output in its products ; this differs from an HMSO click-use licence, whereby a reseller would not only need to obtain a click-use licence for its own re-use or publication of Census data; it would also need to direct its clients to obtain their own click-use licences if those clients themselves wanted to re-use and publish Census data.

The ACD also negotiated an Approved Suppliers scheme with ONS, whereby organisations that could demonstrate that they met agreed standards of data analysis and dissemination may be designated 'Approved Suppliers' of Census value-added products and services. The Census Offices carry out an initial assessment of Approved Suppliers, and carry out annual reviews, so they can ensure that Approved Suppliers provide services of an agreed quality which do not damage the reputation of the Census.

5. OA boundaries – O.S. licence to resupply

As many (most?) delegates at this seminar will be aware, the Output Area (OA) boundaries for the 2001 Census were produced by the Census Offices (ONS, GROS and NISRA). However, they used some Ordnance Survey (O.S.) data in their preparation (e.g. AddressPoint locations, road centre-lines) and therefore needed an agreement with O.S. before these boundaries could be released. (Note this did not apply in Northern Ireland, where OSNI took a different view).

In England and Wales, ONS negotiated an arrangement with O.S. whereby ONS paid O.S. a sum of money to compensate O.S. for future royalties from end-users' own use, subject to ONS and O.S. terms, to cover O.S.' IPR in the data used in designing OA boundaries. Unfortunately, wider O.S. policies excluded from the agreement re-supply of the OA boundaries to third parties as part of commercial value-adding activities. Central Government, Local Government, the Health and Academic sectors – plus end-users in the commercial sector – would not have to pay O.S. royalties to use these boundaries. Value-added resellers (VARs) would have to enter into a licence agreement with O.S., and would need to pay royalties on sale of OA boundaries (whether they charged for them, or not).

Originally, Census VARs (in this case, the Association of Census Distributors, or ACD) expected that terms for reselling, and indeed, payment by ONS to O.S., would have been covered in the overall deal between ONS and O.S. When it became apparent that this was not the case, the ACD approached O.S. Initially, it seemed as though the issue might be resolved simply and quickly. The ACD asked for a one-off payment in lieu of royalties, covered by an agreement similar to that agreed between ONS and the ACD for Census data. The matter dragged on for some months in 2003, and when O.S. came back to ACD it was with a complicated agreement which demanded royalty payments for each OA dataset 'sold' to an end-user (the computation of royalties where less than a total England/Wales dataset was involved would have been tortuous, to say the least). O.S. also insisted that all resellers of Census OA boundaries should sign up to the O.S. Framework Agreement (which would be much more appropriate to resupply of extensive O.S. material on a large scale) – another very complicated document, which required a minimum annual royalty.

The ACD rejected this proposal, for a number of reasons:

- (a). Given that the Census data market had changed radically with the adoption of Census Access – making Census data effectively free at point of use – it seemed logically inappropriate to charge a royalty on the OA boundaries, which could themselves be obtained free from the Census Offices by end-users. ACD members were anyway proposing to pass on OA boundaries to their clients, free of charge.
- (b). The complexity of the licensing arrangements was felt to be out of all proportion to the likely value of the royalties in question. They would have provided a severe challenge to resellers, particularly in terms of calculating royalties (in some cases the resellers would have needed to run the job – for example, a map including OA boundaries – before they were able to calculate the royalties!).
- (c). The need to sign up to the O.S. Framework Agreement (simply to be able to resell OA boundaries, with no other requirement for O.S. products) was felt to be in restraint of trade, and discriminated against smaller 'players', or those to whom mapping was not a major part of their business.

Other correspondence has passed, and meetings have taken place, between the parties, but there has been no resolution of these key issues. Thus it is now three years since the start of this saga, and no agreement exists to resell OA boundaries 'commercially'. The ACD, supported by the MRS CGG, deplors this situation and is planning further initiatives to resolve it. We are urging all parties to arrangements for 2011 Census geography to ensure that this situation does not carry forward into 2011 and beyond.

6. Social Grade output on 2001 Census

The case for deriving approximate Social Grade, in order to produce outputs for market research users, was made over several years via Census BAG meetings. Although there seemed to be an element of doubt over whether this would proceed – due possibly to the parallel development of a new National Statistics Socio Economic Classification – we were very pleased when ONS finally agreed that more than one social classification could be provided for the different sectors of census users.

Although a pilot algorithm had been developed for initial testing on the 1991 Census, this work had to be repeated for the 2001 Census, involving a further considerable amount of work on a tight timescale. All of this work was undertaken by CGG members on a voluntary basis and the algorithm was provided, free of charge, to ONS.

Due to the complexity of the algorithm, the CGG team requested that it be tested by ONS in order to ensure that it had been set up correctly. Although ONS agreed to carry out the test, they refused to supply the evidence that the results matched the data that had been supplied. As a result, the CGG did not identify that there was an issue with coding the 75+ age group until a much later stage – when draft table outlines were issued excluding 75+'s from the base population. The CGG and ONS attempted jointly to correct this gap by introducing additional logic into the algorithm.

During the Census processing phase, the CGG team repeatedly asked ONS to supply results for checking the Social Grade profile however this was refused, on the grounds that no Census results could be released early.

Therefore, when the first Census tables were published and showed issues with the Social Grade results for the 65-74 and 75+ age groups, it was too late to make any changes to the algorithm or tables. The CGG team has had to commission customised tables from ONS in order to examine the issues and provide more meaningful tables for users – involving additional time and cost. These tables have now been produced and are may be obtained from Census Customer Services. Further details of the issues and descriptions of the customised tables are available from the MRS web site:

<http://www.mrs.org.uk/networking/cgg/cggsocialgrade.htm>

We appreciate that Social Grade was a significant development for ONS and are grateful for its introduction, however the outcome would have been more satisfactory if ONS could have taken a more collaborative and helpful position throughout. Ultimately,

however, we believe that the responsibility for shortcomings in the outputs must rest with the CGG team rather than with ONS.

Despite these shortcomings, the Social Grade outputs have been applied by both Census users and distributors, for a variety of market research and analysis purposes.

In view of this, the CGG included a request for Social Grade output from the 2011 Census, based on collection of required occupation, employment and demographic details, as part of its response to the ONS topic consultation last year. We hope that Approximate Social Grade will be included in the 2011 output package and will be willing to work with ONS to redevelop the algorithm – we hope that both sides will learn from the 2001 experience.

7. Other User-Defined Classifications

The Social Grade project was, overall, a positive and beneficial implementation of a segmentation model created by users. We would have liked to see further user-defined models being developed and applied to the Census database.

For example, one could construct a segmentation of households, which would classify each household into a category according to its life-stage and lifestyle. This would be of great benefit to users, for predicting consumer demand more accurately. Users could develop the segmentation model and rules for deploying it on the Census database, if the Census offices could then apply the algorithm – the result would be the segment profile for each area.

The feasibility of creating a household classification has been demonstrated by the CGG and we strongly believe that it should be considered at an early stage for 2011, along with other user-defined models.

The series of ONS area classifications may be viewed as user-defined models - they were built from published Census outputs by users within ONS together with analysts from Leeds University. The Output Area Classification (OAC) would have been of greater benefit to census users in the business sector, if its delivery had been more timely. For 2011, ONS should consider including area classifications as part of the main stream of census outputs in order to bring forward their production.

8. Outputs - Bulk Delivery & Error Management

Customers of the Census Offices, whether resellers or end-users, who require national small-area data are likely to be best served by 'bulk delivery' via CDs/DVDs (or whatever replaces them in future). ONS seemed initially to place reliance on online delivery, which is fine for relatively small areas, but impractical for OA-level data for (say) England/Wales. Thus the bulk delivery requirement was identified relatively late in the dissemination planning process. The actual preferred format for users was csv; again, this had not been anticipated by the Census Offices and indeed, was never successfully

implemented by GROS.

The bulk delivery issues at GROS led to some census distributors having to pay several thousand pounds each for the bulk delivery of Scottish data, in order to meet expected delivery timescales.

ONS had software written to convert output into csv, which was fine; but then, in splitting output by Government Office Region (GOR), caused severe problems for users in loading these data. The position of variables within the same table were different for different GORs - was this problem a feature of Supertable? Whatever its cause, it is well worth bearing in mind for the future, because it causes huge problems for users.

Having recently seen the proposals put forward by Dave Martin, we fully support this approach, and feel it will meet the needs of the ACD and MRS CGG members.

The data releases by the three census offices resulted in supply of over 110 disks to some census distributors. Monitoring and tracking these supplies, together with reported error changes, became highly time-consuming. In reality, some distributors have required only 15-20 of these disks – while ONS must be applauded for providing the 2001 data in additional geographies (e.g. postcode sectors), many of the distributors build their own geographies and could have constructed the higher levels from the data at OA level, combined with necessary geographical look-up tables.

Better consultation on outputs at an early stage would be most welcome for 2011, in order to deliver the files and formats that can be processed most sensibly and efficiently.

Error Management

An associated problem was ONS' mechanism to handle corrections. When the CAS data were released in 2003, there were many errors (some of which are mentioned in 4.6 above). The various processors of the data spotted errors, and advised Customer Services of this; however, it was apparent that Customer Services became submerged under a welter of calls and emails. Subsequently, it became difficult for processors of the data to track responses to their error messages, and there seemed to be poor version management to identify which data release was the latest correction. In hindsight, it is clear that there could have been better systems to record the messages, to keep users advised of progress, and to prioritise the messages (in the sense that there was apparently no mechanism to identify experienced, 'power' users, as against members of the public with less understanding of the data). These issues were addressed following a meeting between the ACD and ONS in November 2003 ; a fast-track system was implemented, and other (former) problems were resolved. However, for 2011 data processing and release, an awareness of these 2001 problems could help with pre-planning and systems.

9. Disclosure Control

We do not intend to revisit previous debates about disclosure control here – the

arguments are familiar to all concerned. However, it is clear that the additional burden of disclosure control had the effect of delaying output release, particularly for small-area data; and that this delay had serious effects on both resellers' and end-users' timescales (leading to both resource issues and financial loss). The 2001 Census initial planning had promised published timetables, which would be met. This promise was not delivered.

We realise the importance of the principle of disclosure control, of course; but we urge that the actual levels of protection be revisited now that data have been published. Is there any evidence of abuse? Is there any evidence of abuse regarding Scottish data? Could there be a review, which then feeds into the planning for the level of disclosure control applying to the 2011 Census?

10. Commissioned tables – the private sector view

When the (excellent) 2001 Census roadshows were presented, the Supercross software was demonstrated, and the standard response to questions from the floor for additional tabular outputs was 'don't worry – you'll be able to get whatever you want through commissioned tables; they will be quick and inexpensive to produce'. The reality has been very different. Several specific examples were given in submissions for this document of requests for commissioned tables, submitted in early 2004, where estimated timescales for delivery from ONS had ranged up to one year! This clearly makes a nonsense of this facility. It is understood that the necessity for tight disclosure control vetting has contributed to the delay. Surely this could be handled automatically (and speedily) by some software solution?

11. Conclusions and look forward

Members of both the ACD and CGG were very pleased (in general) with the quality of 2001 Census output. However, with hindsight we believe improvements were possible.

The most valuable improvement would have been speedier publication of results. We believe that a time frame of data collection in April 2004, to output of detailed CAS statistics in September 2003, is much too long (never mind workplace data in Summer 2004). Given today's data management techniques, it must be possible to publish such data sooner. As a fallback, publication of provisional results at the earliest opportunity may help?

There is no doubt that the requirements of Disclosure Control had a deleterious effect on timings. We remain unconvinced that the degree of disclosure control imposed on the England/Wales data is necessary. We have not heard of any claims to have 'broken' disclosure control on English/Welsh Census data – nor for Scottish data, where the imposition of disclosure control was much 'lighter'. We believe that this issue should be revisited in the light of evidence subsequent to the publication of detailed Census data; and that future decisions should be guided by that evidence.

We have mentioned problems with OA-level output. We do not propose to detail them again here; but believe that the experience gained should be valuable in planning for 2011

dissemination. We hope that the excellent consultation process enjoyed for the 2001 Census will be repeated for the 2011 Census, and that it can be improved right at the end of the process – where output issues should be discussed in detail with experienced users.

We hope that the innovations in the 2001 Census will be carried forward to 2011 – although the trade-off between improvements from the One-Number Census, and the extended timescale, is worth debating. Could the proposed use of administrative records help to both improve results, and speed up the process? We understand that Output Areas will remain as at present (insofar as this is viable – presumably a proportion will need to change?); we assume that 100% coding will continue – and we very much hope that Census Access (or some other mechanism to ensure data are ‘free’ at point of use) will continue.

We look forward to another mutually-beneficial Census consultation process; predicated on meeting user needs.

Peter Sleight, Chairman, ACD
Barry Leventhal, Chairman, MRS CGG

March 2006

P.S. For further information, we would recommend consulting the Geodemographics Knowledge Base (GKB). The GKB - www.geodemographics.org.uk is an international online directory of carefully selected websites on geodemographic and geo-spatial analysis, socio-economic and demographic data sources. The site also includes articles of interest to those working in these areas and publishes an email bulletin.

Association of Census Distributors
Woodlands, Woodlands Close, Holmer Green, Bucks HP15 6QG
Telephone 01494-712371, Fax 01494-714203, Email Peter.Sleight@btinternet.com

MRS Census & Geodemographics Group
The Market Research Society, 15 Northburgh Street, London EC1V 0JR
Telephone 020 8905 2634, Email barryleventhal@uk2.net