

A USER'S VIEW OF THE 2001 CENSUS

Submitted to ONS by :

Association of Census Distributors

MRS Census & Geodemographics Group

February 2005

ACD
Association of Census Distributors



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

A Users' View of the 2001 Census

1. Introduction
2. Purpose
3. Main Issues and Views
 - 3.1 What went well
 - 3.2 What went less well
 - 3.3 Overview commentary
4. More detailed comments
 - 4.1 Consultation
 - 4.2 Collaboration
 - 4.3 Continuity
 - 4.4 Commissioned output
 - 4.5 Consistency between countries
 - 4.6 Bulk delivery
 - 4.7 Corrections procedure
 - 4.8 Disclosure control
 - 4.9 Reseller agreements
 - 4.10 Social Grade
 - 4.11 Additional comments
5. Conclusions and Recommendations

Appendices

Association of Census Distributors - Member Companies

Axiom
Beacon Dodsworth Ltd.
CACI Information Solutions
EuroDirect Database Marketing Ltd.
Experian Business Strategies
MapInfo Limited
SPA Marketing Systems Ltd.

A Users' View of the 2001 Census

Submitted to ONS jointly by:

- (a) The Association of Census Distributors (ACD)
- (b) The MRS Census and Geodemographics Group (CGG)

1. **Introduction**

This report 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 note 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. The reason for our belief that this is important is that 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. Perhaps other user sectors should be encouraged to provide their input?

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.

In order to give a succinct overview of our opinions as to ‘what went well, what went less well’ with the 2001 Census, those opinions are set out below in bullet point form. More detail is given later.

3.1 **What went well**

- * ‘Free data for users at point of use’ – that is, Census Access.
- * 100% coding of all questions and responses
- * The Census consultation process (generally – with some exceptions)
- * Purpose-designed Output Areas

- * Free Output Area boundaries (to users – but see later re.O.S. royalty issues for resellers)
- * Agreement to code Social Grade as an output
- * Implementation of One Number Census to improve results
- * Distribution of data by a variety of electronic means, including NeSS website, and bulk delivery via CDs/DVDs
- * Using postcodes as ‘building bricks’ for OAs, therefore better linkage to postal geography
- * Formatting for desktop GIS use was helpful to large end-users

3.2 **What went less well**

- * Timescales – particularly delays to OA-level data; CAS, univariates, and origin-destination data
- * Quality Control on CAS data output, leading to need for reissues
- * Some problems with bulk delivery – particularly re. formats (and particular problems with GROS and NISRA data)
- * Commissioned tables proving to be non-viable for commercial users, because of quoted delays in processing
- * Lack of consistency between England/Wales, Scotland and Northern Ireland; not so much in terms of ‘special’ questions, but in terms of table layouts (where relevant), common tables where practicable.
- * Although end-users could obtain OA boundaries (free) from the Census Offices, problems with the O.S. agreement, lingering for many months, have frustrated would-be resellers of these data
- * The early promise of Social Grade was sullied by lack of collaboration with MRS experts at latter stages of processing, resulting in flawed results.
- * The absence of income data was frustrating to all user sectors.

3.3 **Overview Commentary**

If some of 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. The fine detail would be inappropriate for this note, but a representative selection of notes from ACD members is attached in anonymised appendices for ONS/GROS/NISRA. We believe this information will be helpful to those charged with dissemination in future.

4. More detailed comments

Many of the points made in the bullet-point sections (3.1 and 3.2) are self-evident, and do not need further elaboration. This is particularly true for those matters where ‘things went well’. However, the remainder of the bullet points need elucidation – which is given below.

4.1 Consultation

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.

4.2 Collaboration

ONS’ agreement to include Social Grade as an output was extremely welcome, and we are grateful for this. However, this initiative went slightly sour in the event. The CGG team went to considerable lengths to develop the algorithm, but at the implementation stage ONS neglected to consult with the CGG team, which had an adverse effect on the final result. This provides a lesson for future collaboration, we feel. See para. 4.10 for more detail of this.

4.3 Continuity

Prior to the 1991 Census, the CGG had recommended the inclusion of a ‘Lifestage’ table, which had been agreed by OPCS, and included in the output as table 53. However, when the first draft of table for the 2001 Census was produced by OPCS, the ‘Lifestage’ table had disappeared – and many other former tables had changed. We were told that the new outputs had been designed by a new team! This may be inevitable, given the ten-year gap; but depending on the content of the Census, previous tables would be a good start-point for development of new ones. The previous learning should be retained, not

abandoned, surely.

4.4 Commissioned Output

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?

4.5 Consistency between Countries

There is no doubt that the lack of coordination between the countries of the U.K. caused problems for processors of 2001 Census data. This ranged through a variety of issues; differential disclosure control, data output format, different numbering of similar tables – the details can be seen in the appendices. A good example of inconsistency may be seen in the religion question, where it seems that in Scotland, the interpretation of this question elicited responses relating to active practice of a religion, whereas in England/Wales, the interpretation was apparently taken as what religion the respondents considered themselves to be (hence the occurrence of ‘no religion’ is about double in Scotland versus England/Wales). We realise that devolution has brought independence of decision making, and that there is a demand within devolved administrations for some differences in questions; but notwithstanding this, we would hope for better coordination in future in order to meet the needs of users and processors of U.K.-wide statistics.

4.6 Bulk Delivery

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). 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 (more detail of this, and associated problems, may be found in the appendices). 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.

4.7 Corrections procedures

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.

4.8 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). Disclosure Control has also severely compromised some planned output, notably SARs and some workplace statistics. 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?

4.9 Reseller Agreements

This is, of course, a particular issue relating only to would-be value added resellers of Census data (VARs). Since the formation of the ACA (now ACD) in early 1994, the organisation has enjoyed a good working relationship with OPCS/ONS. Discussions about a new 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 we 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. This went through a number of iterations, and was virtually complete in early 2001, but was not finalised for a number of months in the event (actually, after Census day). In planning for 2011, we believe it will be important to have an agreement in place well before Census Day, so that prospective customers' queries may be answered. This is actually very important to VARs; clients find it hard to believe that reselling terms and conditions cannot be established in advance of the taking of the Census.

The insistence of Ordnance Survey on demanding royalties from resellers of OA boundary data has caused ongoing problems for the ACD. While the ACD continues to address this with O.S. and ODPM, this issue might have been foreseen by ONS – indeed, the ACA (predecessor of the ACD) raised this issue in 1997. We believe it will be crucial to resolve this issue well before agreements for 2011 and onwards are formulated.

4.10 Social Grade

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.

A pilot algorithm had been developed for initial testing on the 1991 Census, however this development 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. At the time of writing, we are waiting for these tables to go through the 'commissioned outputs' process (so far, this has taken one year).

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.

4.11 Additional Comments

A few miscellaneous comments from CGG members:

- (a) An end-user who did his own processing of Census data compliments the Census Customer Services team at ONS for their excellent service; also ONS for population-weighted centroids of OAs, which added attribute information and increased accuracy of analysis. He also commented that Super Output Areas could usefully be provided with population-weighted centroids to increase their usefulness.
- (b) Another comment which arose several times related to the fact that English/Welsh OA boundaries were defined differently from Scottish OA boundaries. Whereas Scottish OA boundaries were cut to high mean watermark, English/Welsh OA boundaries were not. The effect of this is that density calculations can be accurately made for Scotland, but not for England/Wales (unless the OA boundaries are 'doctored' in England/Wales).

5. Conclusions and recommendations

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 2001, 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

February 2005.

Appendix A
Company A - Reply regarding ACD evaluation of the 2001 Census

1. Planning

(a) Distributor agreements

- It is felt that the end result here is satisfactory in the light of current Census dissemination policy, but the process of getting to the current position was onerous. The change in dissemination policy in 2002 effectively meant that most of the lengthy representations and negotiations up to that point were time wasted for all parties. In terms of timing, the fact that distributor arrangements were not in place at the point when the Census was undertaken (when the most client interest arises) was disappointing and should be the goal for the next Census. In terms of process, the mechanism of the ACA (ACD) negotiating as a single entity was undoubtedly helpful to all parties.

(b) Third-party arrangements (eg OS)

- The only significant 3rd party here is the OS. There is still no agreement in place in May 2004. The ONS rather than the ACD undertook to manage the process of negotiation with the OS. Whilst best efforts were undoubtedly made, the ACD is now having to negotiate direct – even this has taken time and there is no resolution yet. The amount of management effort and time spent on this comparatively trivial part of the output has been completely disproportionate.

(c) Mechanism for informing/involving all parties in Census planning

- Company A involvement in this was mostly via the CGG.
- It is felt that the mechanism was generally good with plenty of opportunity for input and feedback when appropriate. The mechanism and communication channels allowed several key advances to be made compared to the 1991 Census such as the production of ‘One Number’ Census results, the modelling of Social Class, and the inclusion of a question about Religion.
- The communication channels also allowed concern to be voiced regarding other issues such as the use of ‘postback’ for form return, the lack of a question on income, and the originally proposed method of ‘small count adjustment’ for disclosure control. Whilst feedback/lobbying did not affect the first two of these issues it did result in an amendment to the ‘small count adjustment’ algorithm which was far more acceptable than the originally proposed method.

(d) Consultations on form of output

- Again, it is felt that this was generally good.
- However, two areas where it is felt that the consultation and communication regarding output was poor were Social Class and Workplace statistics.

Social Class

- It is felt that decisions regarding the application of the Social Class algorithm and output were made by ONS without adequate consultation regarding :-
 - (a) the data to be captured
 - (b) the design of tables to be output, and
 - (c) the production of statistics without adequate opportunity for QA and feedback by the originators of the algorithmThis resulted in some last-minute adjustments to the algorithm being required and the publication of tables with inaccurate results for older people.

Workplace

- Whilst it is accepted that circumstances can change it is felt that it is very unfortunate that ONS concerns regarding confidentiality and ‘Disclosure Control’ for Workplace Statistics did not come to light until very soon before the expected publication date (Note :- It is believed that this publication date was already a later date than that originally quoted). This concern led to a further lengthy delay in the publication of the statistics and the decision not to publish some tables where, from my understanding, ‘final specifications’ had previously been published. It would obviously have been better if these concerns had been raised as part of the original planning and design of the Census or even as part of the subsequent assessment and evaluation of various ‘Disclosure Control’ methods that took place regarding residential statistics. It would be useful to understand if this was an oversight on the part of the ONS or if changing circumstances led to the decision to review ‘Disclosure Control’ methods for workplace statistics.

(e) Consistency between countries

- Company A feels that the opportunity and freedom for each Census Office to design questions and produce output that is different to other parts of the UK in order to meet localised needs is a good aspect of the Census. (*It should also not be forgotten that this provides an opportunity for Census Distributors to ‘add-value’ to the Census results by collating UK-wide datasets*).
- However, there are areas where greater consistency between countries would be helpful as follows :-
 - Disclosure Control methodology – it is difficult to explain to people why different levels of ‘Disclosure Control’ are felt to be adequate depending upon whether, for example, you live in England or Scotland. It would be interesting to know if the ONS would reconsider their approach to ‘Disclosure Control’ if in the years preceding the 2011 Census there are no instances of breaches of confidentiality in Scotland.
 - Timing of outputs – There was a gap of approximately 5-6 months between the release of the majority of CAS tables for Scotland (March 2003) and those for England & Wales and N.Ireland. It would be very useful in future if the timing of the same output from different Census offices could be closer thus allowing the quicker construction of GB or UK datasets.
 - Data Output Format – It is believed that ONS/GRO(S)/NISRA were using different versions of SCROL to produce their output. One consequence of this was that programs written by ONS to convert SCROL output into a more user-friendly format (standard CSV output) could not be used by GRO(S) or

NISRA. It would obviously be beneficial in future if there was greater consistency in output formats between countries.

2. Output Delivery

(a) Bulk delivery – pre-planning

- It is felt that the various meetings/workshops/seminars held by ONS regarding output formats etc were very useful. It would have been helpful if GRO(S) and NISRA could have played a greater part in these events to help the consultation process to be undertaken on a UK-wide, rather than separate country, basis.
- ONS were also very helpful in accepting and acting upon the issue of the poor format of SCROL output. They allocated resource to write a customised set of programs to convert SCROL output to a much more usable ‘standard CSV format’.
- One aspect of CAS delivery by ONS that Company A were not aware of until the last minute was the delivery of CAS tables split into separate files by Government Office Region (GOR). It is felt that this led to several problems that are detailed in 2(d) below – ‘Quality Control’.

(b) SCROL problems

- Issues relating to the poor format of SCROL output (ie it is very difficult to use) and the use of different versions of SCROL by the different Census Offices are covered in sections 1(c) and 2(a) above.

(c) Opportunity for beta-testing

- The amount of beta-testing data that Company A received from the various Census Offices varied as follows :-
 - ONS :- Two samples of beta-test data were received (featuring mocked-up or 1991 Census data). One sample in ‘SCROL CSV format’ and one sample in ‘Standard CSV’ format. The supply of this beta-test data was very helpful and useful to us in order to prepare for receipt of the actual data.
 - GRO(S) :- No beta-test data was received. However, this did not prove problematic as the release date of the GRO(S) CAS data (March 2003) was much earlier than that for the other Census Offices and few problems were encountered in loading this data.
 - NISRA :- No beta-test data was received. This, combined with the timescale for delivery of data for N.Ireland and the changing of report layouts at the very last minute did cause some problems when loading the N.Ireland data.

(d) Quality Control

- The nature of some of the errors that Company A have found with the CAS data supplied by ONS is of great concern and indicates that QA procedures need to be improved next time around.
- Examples of the types of errors that we have found are :-
 - The position of variables within the same table being different for the separate files supplied by GOR (eg variable X in a table may appear in, say, the 10th column for one GOR but appear in, say, the 11th column for another GOR)
 - Some variables not being populated at all.
- One aspect of data delivery that made it difficult for Company A to undertake its own checks to confirm that the supplied data had been loaded correctly was the fact that no checking statistics (eg variable totals for England & Wales or by each GOR) were supplied by ONS. The ‘Disclosure Protection’ method used for England & Wales exacerbated this problem as it meant that we could not cross-check variable totals against a National Report to see if there was an exact match. In contrast, this could be done for data received for Scotland given the different ‘Disclosure Protection’ methods undertaken by GRO(S).
- As well as not receiving national or GOR checking statistics ourselves we believe that ONS may have been unable to produce these statistics for themselves. If this is the case then it is likely to have played a major part in some of the types of error described above not being spotted by the ONS.
- Given the above we think that it is vital the ONS QA procedures are strengthened next time around and that increased resource is allocated to both plan and undertake these procedures. In particular, the production and supply of national and regional variable totals based on the CAS (or equivalent) should be given a very high priority next time. We think that the ONS should also consider undertaking more ‘spot-checks’ of the data and look at automating some checking procedures to produce suitable ‘exception reports’. For example, an automatic comparison of the penetration rates of variables between GORs followed by manual inspection of outliers would help to identify cases where a variable appeared in different column positions by GOR. It is hoped that the automation of checking procedures would keep human resource requirements down to an acceptable level.

(e) Other delivery issues

There are three other main delivery issues that we would like to mention:-

- Specifications of tables
- Timetabling of Output
- Customer Support

Specifications of tables

- It was a good idea to have table specifications made available via the internet before the release of the Census results.

- Unfortunately, two things limited the usefulness of these specifications :-

(i) *Poor Version Control :-*

Sometime before the publication of the Census results the table specifications were labelled as ‘final specifications’. This was not the case as several changes were subsequently made to them. This situation was made worse by the fact that there seemed to be little or no version control of the table specifications PDF file. This made it very difficult (if not impossible) for the user to determine which table layouts had changed (and what these changes were) from one download of the PDF file to the next.

Probably the most extreme examples of poor version control occurred in N.Ireland. Firstly, the tables for N.Ireland Key Statistics that were released contained several differences to the PDF file table specifications still available on the day of release. Secondly, the PDF file containing CAS table definitions was withdrawn from the NISRA website and not replaced by an updated version. The result of this was that it was impossible to know N.Ireland CAS table final definitions until the data actually arrived.

The publication of table specifications believed to be ‘final specifications’ which were then subsequently altered led to a certain amount of wasted planning and programming at Company A.

(ii) *Format of specifications :-*

It would have been useful if the table specifications could have been made available in other formats (eg EXCEL, Word) as well as PDF. This is because other formats would allow users to cut and paste required table definitions into their ‘data load programs’ etc – this cannot be done with PDF downloads.

Timetabling of Output

- GRO(S) (Scotland) were excellent at keeping to initially published release dates.
- ONS, for various reasons, obviously had to revise and republish release dates on several occasions. ONS were generally good at informing users (via Census News) of any revisions to release dates. They were also good at either keeping to these revised release dates or informing users with sufficient notice that there had been further revisions to them.
- The release dates that NISRA published in its Output Prospectus were very vague – for many products the release date was not quoted and, when it was quoted, it frequently only stated a particular season within a year. Company A found that it was normally necessary to phone the NISRA helpdesk in order to obtain release dates of sufficient precision to add the planning & scheduling of our projects.

Customer Support

- The Customer Support, especially Helpdesk Support, provided by all three Census Offices has generally been good. Census Office staff have generally been very helpful in logging and responding to queries.
- The main issues with Customer Support have been as follows :-

(i) *Overloaded after initially CAS release*

It appeared that the ONS Helpdesk was overloaded following the initial release of the CAS. It is not known whether this was due to a higher than expected number of issues with the CAS, inaccurate capacity planning, budget limitations or a combination of all three of these factors.

(ii) *Lack of query tracking*

It appears that the ONS Helpdesk staff have difficulty retrieving information about queries and tracking their status if subsequent calls are made to them. If Helpdesk staff had instant access to a database which held such information this could help to make the service more efficient.

3. **Other Comments**

The comments in sections 1 and 2 above cover the aspects of Census Offices' performance as listed in Peter Sleight's note to the ACD of 28th April 2004. It is felt that the comments provided above tend to emphasise problems associated with the 2001 Census and do not give enough prominence to its successes.

In this section the key strengths/successes and weaknesses/failures of the 2001 Census from Company A's point of view are therefore listed to help a balanced view to be obtained.

Key Strengths/Successes

There is no doubt that there have been some key strengths/successes of the 2001 Census, especially in terms of innovations compared to some aspects of the 1991 Census.

Company A feels that these have been :-

- ***Consultation by ONS/GRO(S)/NISRA***

As stated above, it is felt that the consultation process has generally been very good and that ONS/GRO(S)/NISRA have been willing to listen to users comments and generally act on them if it has been practical to do so. The consultation and planning process resulted in some ambitious plans for the 2001 Census that have mostly been successful.

- ***Introduction of 'Output Areas'***

There is no doubt that the introduction of Output Areas and their benefits when compared to Enumeration Districts (eg smaller, more homogenous etc) has been a major step forward from the 1991 Census. The adoption of Output Areas as a key

geography by ONS (eg their use within Neighbourhood Statistics) will bring further benefits over time.

- ***100% coding of results***

This is a significant improvement over the 10% coding of the 'hard to code' variables from the 1991 census.

- ***'One Number Census' Methodology and Output***

Whilst there has been criticism of some of the results (especially relating to inner-city areas) and some aspects of the methodology it is our understanding that the application of the 'One Number Census' Methodology has been a 'world first' and has generally been very successful. There is no doubt that the publication of Census results that relate to the whole population by allowing for differences in under-enumeration between areas is a major benefit. It is hoped that the ONC methodology can be built on and improved for the 2011 Census

- ***New output such as 'Social Class', NS-SeC and 'Religion'***

The new questions and output (compared to 1991) are proving very useful.

- ***Distribution mechanisms for results***

The delivery of a subset of results and commentary via the internet has been successful as has the use of DVDs/CDs as distribution mechanisms for bulk data.

Key Weaknesses/Failures

Company A's views on problems associated with the 2001 Census tend to have been covered in sections 1 and 2 above. Nevertheless, it is felt to be worthwhile highlighting (and sometime reiterating) below what we perceive the **key** problems to have been :-

- ***Sign-off of Output Table Specifications***

The text associated with PDF files of table specifications published early on via the ONS/GRO(S)/NISRA websites led Company A to believe that these versions of table specifications were final specifications. This did not turn out to be the case with further versions of table specifications subsequently being published by ONS and GRO(S). The situation for NISRA was even worse with the final specifications for tables only being known by examining the data after it had arrived. These changes from 'final specifications' led to a certain amount of wasted planning and programming by Company A as well as more work being needed to load tables when the data arrived. It is hoped that next time either (a) the fact that published specifications may change in the future will be given more prominence on the Census Offices' websites, or (b) preferably, specifications presented as 'final specifications' will actually be final. It is also preferable that final specifications (which are final) are published in far as advance as possible before the data actually arrives in order to allow adequate planning.

- ***Version Control of documentation (especially Output Table Specifications)***

When documentation was changed (especially Output Table Specifications) then the version control to log and highlight changes tended to be poor or non-existent. It would obviously be helpful if improved version control could be applied in future

- ***Changing (and delayed) release schedule***

The delays in the initially announced release schedule have definitely been frustrating for users of the data. It is accepted that some delays are always likely to occur. However, the length of some of the ONS delays and some of the reasons for them have surely been avoidable. In particular, delays caused by changes to ‘Disclosure Control’ methods could surely have been avoided or reduced with better planning – especially for Workplace data where the issue of ‘Disclosure Control’ seems to have been considered very late in the day.

- ***QA of data before release***

As mentioned above, there have definitely been some areas of QA that could have been better – especially for ONS. In order to improve QA next time it is hoped that the Census Offices will place greater emphasis on planning QA procedures, increase the resource allocated to this if required, and consider how QA procedures can be automated to highlight possible errors and thus make more efficient use of staff.

- ***Problems associated with application of Social Class algorithm***

The agreement by ONS to run the Social Class algorithm and include results on estimated Social Class within the standard Census Output was a key innovation that was welcomed and supported by the business community. It is unfortunate that, for various reasons detailed above, the final results in the Social Class tables have not been as accurate as one would like. It is hoped that ONS will be able to undertake the necessary work to re-run and re-publish the Social Class tables in order to make them more accurate and usable. This will allow the work that has been put into this project to be fully exploited.

ONS Appendix B

Subject: Company B Census 2001 Data Processing Review”

Date: 18th May 2004

Version: 0.1

“We recommend ONS, GROS and NISRA undertake a review of the methodology used to deliver the 2001 Census data and implement changes in time for the 2011 Census.”

1. ONS website “Known Error Log” was not updated on a regular basis
2. Known errors were not communicated to all of the user population
3. Metadata on the ONS website did not relate to actual data
4. Data taking multiple efforts to get right
 - a. Tables
 - b. Geographies
 - c. Population Headcounts
5. GROS and NISRA released ASCII data in a unnecessarily more complex format (Excel/CSV layout) than ONS data (fixed length record ASCII) – indicated poor technical understanding (Excel could not handle the volume of records and the format being asked of it)
6. More emphasis appears to have been put into the SUPERTABLE release than the quality of the ASCII data
7. No effective internal QC was undertaken before ONS released the ASCII data
8. There was no cross-agency consistency in the data gathered or the methodology of release
9. Time taken for the ONS to gather, process and release the data was too long – numerous deadlines came and went
10. The ONS did not deliver a “One Number Census (ONS)”
 - a. Totals derived from CAS data were inconsistent with that derived from Standard data due to disclosure controls (GROS and NISRA delivered a “ONC”)
 - b. ONS did not release any national totals based on CAS data – impossible to then QC CAS level data after amending anomalies
11. Metadata was made too complex by all agencies – ONS could have released one table with all metadata in it; GROS and NISRA could have released more easily accessible metadata (Excel/CSV layout was too complex)
12. GROS should have released data in national tables – the Local Authority level it was released at was probably more complex to derive than a national one would have been
13. ONS should have released data in national tables – the Regional level it was released at was probably more complex to derive than a national one would have been

14. ONS relied on the SUPERTABLE software too much – it's limitations were lived with and not addressed:
 - a. SUPERTABLE was unable to process 'large' datasets – national table were therefore unable to be processed – this is technically unacceptable in this day and age
 - b. SUPERTABLE was unable to consistently output data:
 - i. The order of variables was randomly mixed up across outputs
 - ii. Metadata fields were truncated
 - iii. Variables were output for one region but not another
 - iv. Zero suppressed OAs dealt with inconsistently
15. Naming conventions were inconsistent throughout the ONS metadata (e.g. Table KS06 in one piece of metadata, KS006 in another)
16. Naming of variables inconsistent across agencies
17. The anomalies reported above resulted in:
 - a. Data being processed by multiple clients
 - b. Potential for each client to produce conflicting figures (when a "ONC" had been promised for years by ONS)
 - c. Time and money being wasted by agencies and clients
 - d. Poor PR for Census Offices
18. Recommendations to all Census Offices
 - a. Don't be dependent on 'black box' systems
 - b. Be more technically literate
 - c. Be more data literate
 - d. Involve independent 3rd parties in QC process
 - e. Speak to clients

APPENDIX C : COMPANY C - 2001 CENSUS REVIEW

1. Planning

- (a) Distributor agreements – need to be finalised in plenty of time

All agreements need to be in place well before and data release. We did achieve this and the agreement is well structured and we are pleased with it

- (b) Third-party data arrangements (e.g. O.S.)

As a) these should be in place well before any release of data. The position with OS is ridiculous and could have been avoided, after all we did warn them.

- (c) Mechanism for informing/involving all parties in Census planning (e.g., via BAG, CGG etc.)

The roadshows were an excellent idea where important issues were aired. It also gave a platform for discussion. I think the Census Agencies meetings were also very useful to ensure our own specific input and influence.

- (d) Consultations on form of output (as above)

Excellent documentation that gave confidence on good well thought out output. It offered lots of opportunity to discuss.

- (e) Consistency between countries (or not! ONS, GROS, NISRA)

All 3 organisations seem to forget that we are still 1 country. A number of frustrating/irritating inconsistencies between countries became apparent in the consultation and even more so in the eventual output.

e.g. (A few examples)

Key Statistics: supplied as counts by ONS (documentation says percentages), percentages by GROS, and both counts and percentages by NISRA.

Table numbers and structures are inconsistent, especially between Northern Ireland and the rest of the UK. CAS01: all variables after the first cell do not match. CAS049 (Dwelling type and accommodation type by tenure), the Northern Ireland equivalent is CAS353.

Age bandings in CAS from Northern Ireland compared to GROS and ONS.

Ethnicity: "Mixed ethnic group" definition in KS06 differs between countries.

Religion: variables inconsistent in KS07 (KS07A for Northern Ireland). Northern Ireland splits Christian into 5 categories, and puts everyone else in "Other", whereas the rest of the UK has Christian as one category and has 6 other categories.

Inconsistent approach to the derivation of Output Areas – different methodologies and rules.

2. Output Delivery

- (a) Bulk delivery – pre-planning

This was just not thought about before it was almost too late. Only ONS eventually met the challenge and Scotland and Northern Ireland just ignored the problem.

ONS provided csv format but then that was received with a price: poor control on precise output e.g. cell transposition. National level tables would also have been preferable rather than having 10 (regional level) input files for each output table.

Scrol dumps from GROS and the data from NISRA were horrible and a nightmare to process.

It totally defeated the excellent consultation process and production of all the tables etc for any forward planning by recipients of bulk data. If the delivered data does not match the published proposed format the information is useless.

(b) SCROL problems

Excellent for using as a checking mechanism but SuperTABLE is a wholly inappropriate piece of software for bulk delivery. Extremely cumbersome to extract data from. Great if you are just doing simple area analysis.

Bulk delivery from GROS was just a dump of SCROL and it had to be paid for. We ended up having a third party process for us, The GLA, who did an excellent job and delivered in the correct format. Data dumps from SCROL/SuperTABLE were particularly bad for CAS tables where geography and data headings were mixed up together.

(c) Opportunity for beta testing? (issues?)

There would appear to be some need for beta testing. I would expect due to the restrictions highlighted by ONS this would need to be an internal team. Our recommendation that a team divorced from the processing and independent of the census group should provide objective critique of output. I think we all find this process invaluable in raising issues prior to a full delivery by asking "daft" and awkward questions.

(d) Quality Control generally

Lots of upfront documentation but too many mistakes at delivery. What was received did not match what was expected (e.g. Key Statistics coming as a mixture of counts and percentages). There appeared to be little control on bulk delivery.

Updates were provided with little information on what was being replaced and precisely what had been fixed.

Once raised, a Fast Track system was implemented, which was good but a bit on the late side.

List of errors on the website quite often did not reflect issues raised by us.

(e) Other delivery issues, e.g.

- dependency on 'black box' software
- organisation to handle output
- timetabling of output
- version control of re-releases
- access to information about details of changes

Trying to keep up with the timetable was a skill in itself. Biggest problem was managing expectations in the business (and therefore clients) about when data was going to be available and when we might be able to deliver the new products. The timetable was changed quite regularly.

Disclosure control by ONS was a major problem. Undermined the whole output in the wider community.

Having to keep in contact with and manage delivery from 3 organisations could have been done better. This just trebled the workload, really needed one contact to provide national data sets. Maybe even join them up before delivery.

There were a lot of excellent data sets. In particular the Postcode Headcounts and the Output Area to Postcode lookup were excellent developments.

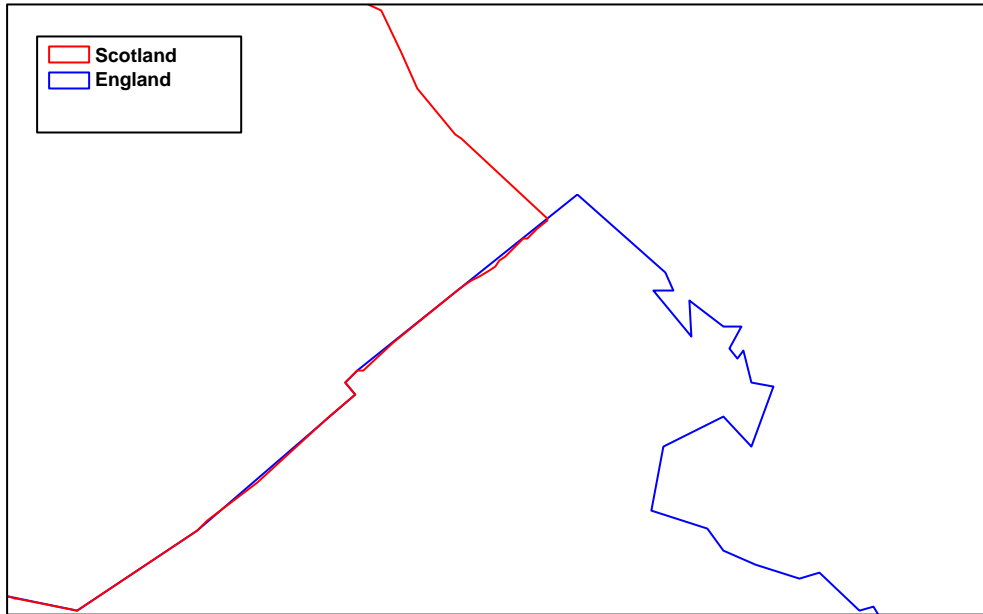
(f) Output Area Boundary Data

Output Area Boundary data in Mapinfo format from GROS and NISRA was OK.

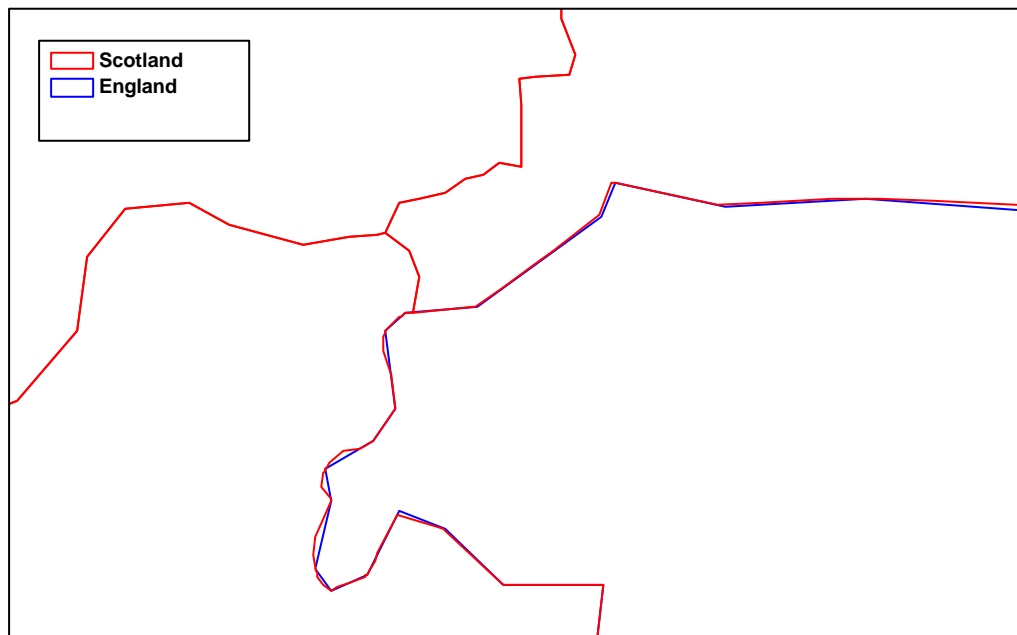
ONS Output Area Boundary data in Mapinfo format was extremely problematic. A lot of problems with complex bowties and slithers, which made it unusable under certain functions in Mapinfo, in particular combine objects. Cleaned considerably but still a lot of problems. Required a huge amount of work to make useable.

Scottish and English borders do not match at all. Scotland is slightly off to the West next to England, see below.

Scottish and English Border from delivered Output Area Boundaries
1. The offset between England And Scotland



The Badly matching border



ONS Appendix D

Views about ONS' 2001 Census performance.

1. Emphasise that providing the data by question within region was very laborious to use. I think that they should provide the data as Access files in the future (Excel only goes to 60,000 records) to avoid the errors associated with joining the regional questions together and to make it more convenient. It may be that in providing the CSE files ONS went through Excel and were therefore very limited in what they could do - but this all caused a lot of pain (but not to LADS of course!).
2. Labels, which have leading 00, can easily be seen by the software to be numeric format instead of the string that it should be and get reduced just one 0. An example is the county label and the settlement labels in Scotland. This has happened to me on a number of occasions and is extremely inconvenient and a big fiddle to sort out. I think ONS should avoid such labels. It may be that the string nature is held within in the initial file format but it can get lost when importing and exporting a file between different software. For example, I was regularly moving data between Access and Excel and SPSS and putting it into various formats in the process including DBF format for use in ARCVIEW.
3. Disclosure problems prevented a number of univariate tables being published in workplace - this has to be silly.
4. Grinding an axe - disclosure control should be variably applied depending upon the density of the potential disclosing target. For example, a two way table of ethnicity and occupation in rural Somerset may well be disclosive whereas it would not be in Southall. People will be interested in this table in Southall but not care about it in rural Somerset. If all the areas where disclosure was possible were damaged such that the table of the sum of these areas across the country was correct, then the whole database would be correct. Surely they should be able to get someone to work out how to do this.

Martin Callingham January 2005.