AN ON-LINE, SECURE AND INFINITELY FLEXIBLE DATA BASE SYSTEM FOR THE NATIONAL POPULATION CENSUS

D.W. Rhind, E. Hayes - Hall, H.M. Mounsey[1] and S. Openshaw[2]

1 : Department of Geography, Birkbeck College, University of London, 7-15 Gresse Street, London W1P 1PA, UK
2 : Department of Geography, The University, Newcastle, UK

ABSTRACT

This paper describes a prototype on - line data base system to handle the 20 million or more records which will arise from responses to the individual questionnaires for the next Census of Population in Britain in 1991. The work is funded by the UK census agencies and by the Economic and Social Research Council. Unlike conventional census output, it is predicated upon producing the necessary results 'on demand' for variables and areas specified by users in a variety of ways over national telecomms networks. The methods of achieving this have involved use of two quite different computer systems, one being primarily a hardware - based solution and the other a more general, software - based approach. These two approaches are described, together with the means of maintaining confidentiality in the 'raw' data (as required by statute) and the implications of operation of such a system for value - added census services.

INTRODUCTION

Statistics from the decennial Census of Population form possibly the single most widely used data source in Britain (DoE 1987). Traditionally, these statistics have - like those from most other censuses - been produced as area aggregate statistics in count or cross - tabulated form (Redfern 1987, Dewdney and Rhind 1986). These have described the overall characteristics of the people or the households in each areal unit and their form has had to be planned long in advance of the census itself (see various chapters in Rhind 1983). In Britain, the Small Area Statistics (SAS) are the best example of such a pre-defined set of tables; inevitably, these reflect compromises after discussions with many potential users of the statistics (Denham and Rhind 1983). In contrast, unanticipated requirements are met by special tabulations, charged and produced to the specification of individual customers, largely after the production of the standard census statistics. Distribution of all these statistics has hitherto been in paper form or on magnetic media.

Two census - taking agencies exist for mainland Britain : these are the Office of Population Censuses and Surveys (OPCS) in England and Wales and the General Registrar's Office for Scotland (GRO(S)). Collectively, they have long recognised that this dual form of output is scarcely ideal. Aided by increasing densities and diminishing costs of computer storage, they have sought to provide a wider range of information in the standard products. The 1971 Census SAS for England, for instance, consisted of only 1,571 cells of cross-tabulated information derived from 30 questions whereas the 1981 equivalent consisted of 4,400 cells from only 21 questions. It has recently been proposed that the standard 8abulations for the 1991 Census should contain about 6,000 cells (OPCS 1988). This 'shot gun' approach has had a price; as analyses of Census data used by active user organisations such as Tyne and Wear County Council have indicated, any one organisation may never use many of the cells and some users have been unable to afford or wait for the special tabulations. Finally, the volume of cross-tabulated summary statistics has come increasingly close to those of the 'raw' data : the most detailed 1991 SAS data, suitably compressed, are likely to occupy about 750Mb whilst bit-compressed 'raw' data could amount to between 600Mb and 1Gb.

This approach to census data processing stemmed from 1960's and 1970's technologies; it led directly to the advent of a single package commissioned by a consortium of (mostly local government) users to handle the 1981 SAS; the result, called SASPAC, ran on many different ranges of computers and operating systems and was arguably the most portable product of its day. It was responsible for introducing perhaps 2,000 individuals to census analysis (Rhind 1984). Despite SASPAC's success, the software was a creature of its time, being necessarily tailored to virtually the lowest common denominator of computing availability. The institutional situation at that time was typified by the sole availability of stand-alone mainframe computers run by centralised administrations, by batch access, by relatively restricted and 'unfriendly' software and by a punched card record orientation.

This world has changed totally. Desk top micros selling for around $1,000 now have far more data storage and processing capabilities than many of the mainframes in use in many local authorities in 1981. In particular, the entire SAS data for all 130,000 Enumeration Districts (EDs - the approximate equivalent of Collector's Districts) in Britain could, for instance, now be stored on a CD-ROM of the type which the UK Post Office are now selling: their price of $7,000 for hardware reader and disc containing details of 1.5 million unit postcodes indicates the fundamental nature of the change in the last six years - particularly since the price level set reflects the UK government's desire to maximise its financial return on data collated under its auspices. In November 1988, the census offices revealed plans by a number of commercial firms to distribute 1981 SAS on such media and many other national (e.g. the Swedish Land Survey) and international agencies (e.g. World Data Centre A in Boulder, CO.) and commercial agencies are already distributing data by these methods.

Moreover - in universities and some businesses at least - the use of computer networking to move files or to access specialist software on remote computers is now an everyday activity. Distributed databases and distributed computing power is becoming a reality and recognised as such - the Computer Board for the Universities in the UK (which is responsible for approving all central planning of computing facilities in all British universities) argued this to be the most far-reaching change in computing in its three yearly report published in December 1988. As another example, the feasibility of running a three site US National Centre for Geographic Information and Analysis and an eight member, 13 site set of Regional Research Laboratories in the UK is substantially dependent upon routine use of electronic mail and remote access to software and data bases.

There is every reason to believe that these technical and technological developments will continue and will spread to other sectors. Computing facilities in 1991, then, will be unrecognisable as compared to those of a decade earlier.

SOME NEW PROPOSALS FOR CENSUS DATA HANDLING

Given all this, Rhind (1985) argued for an equally fundamental review of census strategy. He outlined two options, the more radical of which saw the census mainly as a calibration tool with a wider range of data being assembled from various other sources and linked together. The second option was one in which OPCS/GRO(S) would produce only the most basic summary statistics and distribute these, say, on floppy disc for PCs or their 1991/2 equivalent; it is certain that much competitive software will be available to read, analyse statistically, tabulate and map a set of standard statistics for each areal unit. In tandem with this, an on-line service was proposed to be provided by OPCS/GRO(S) for querying a data base of unaggregated Census records, from which users could extract any desired aggregate information - for whatever area(s) or groups of people and cross-tabulated as required - unless this was liable to disclose details of an identifiable individual or household or lead to unreasonable intrusions into privacy. It was further proposed that this would be achieved through building up an Intelligent Knowledge Based System incorporating whatever rules OPCS/GRO(S) already applied and whatever others were found to be necessary.

Following discussions, OPCS and GRO(S) provided a contribution of £11,000 towards the funding of a project to investigate the feasibility of such a system and preliminary work began in late 1986. The funding by OPCS/GRO(S) of course, does not imply any commitment to introduce such an on-line system, should it prove feasible. Following a further grant of £22,000 from the ESRC from November 1987, work accelerated. By April 1988, progress was sufficient for an initial demonstration of the system and for a clause referring to possible use of such a tool to be included in the White Paper on the Population Census (HMSO 1988 para 53).

PROJECT OBJECTIVES

The objectives of the research project were:

(i) to create a prototype on-line storage and retrieval system capable of handling Census questionnaire returns from the 1991 Census. This would permit :

- users to specify which counts or cross - tabulations they require at any particular level of aggregation of variables and / or geographical area. So far as geography is concerned, we have assumed that at least three ways of defining areas of interest are required : using the standard census nested area hierachy (ED/Ward/District/County/Region/Country) and any variant of it (e.g. ED/Health District/Health Region) as 'building blocks'; equivalent hierachical and ad-hoc combinations of unit postcodes; and use of National Grid References. This pre-supposes multiple indexing. So far as census variables are concerned, we assume all possible combinations of responses to census questions within one return should be feasible.

- users to obtain an estimate of cost prior to the execution of the job,

- confidentiality of the 'raw' data to be maintained,

- print out of all standard SAS and other tabulations as well as the 'special ones'.

(ii) to create 'pseudo 1981 Census' questionnaire responses for sufficient areas to test the system adequately.

(iii) to demonstrate the system described in (i) on data derived from (ii).

(iv) to document the results, including a summary of the aggregation and data release rules finally agreed with OPCS and GRO(S), the level of 'noise' required to maintain confidentiality, the results of user trials and predicted implications of introducting such a system into operational use.

For reasons which are described below, the initial approach was developed only to the stage where it demonstrated that many of the project objectives could be met. The approach was based upon the creation of a superstructure to the ICL INDEPOL system : this exploits the high security and speed of access provided by ICL hardware, the Content Addressable File Store or CAFS (Carmichael 1986, Wise and Pellett 1988). ICL facilities have been in use by OPCS and GRO(S) for 30 years and CAFS is a standard facility in all ICL 3900 mainframes. However, following a review of OPCS' IT strategy by a firm of management consultants, it was decided to base future strategies on an initial choice of a software environment, rather than selecting the hardware and then obtaining or creating the necessary software. Model 204, a relational database with such features as the Structured Query Language (SQL), was selected to form the main part of the new OPCS software environment. Since there is no current implementation of Model 204 on ICL equipment, this decision by the census offices - taken for much broader considerations than any one particular tool (especially one being developed in a research study) - necessitated an equipment procurement which might well preclude use of our ICL - specific system.

Faced with this, we opted to mimic the facilities already developed on the ICL equipment by re-implementing them (so far as possible) under Model 204. Fortunately, we found one university installation in the country using the software and were kindly allowed access to the software running on the University of Manchester Regional Computer Centre. At the time of writing, OPCS had not finalised which hardware and software would be used for the entirety of the 1991 Census data processing; a mixed ICL/Model 204 configuration seemed a real possibility. For this reason, this paper contains sections on the technical aspects and a preliminary assessment of the relative merits of each of the two approaches.

THE TEST DATA

The data base required to test and to demonstrate the resulting prototype software was the initial priority in the project. 'Raw' data from recent censuses are not, of course, available because of the statutory constraints embedded in the 1920 Census Act; this precludes the release of data pertaining to identifiable individuals in less than 100 years after its collection. The obvious solution of coding up 1881 or earlier records was rejected because of the resource implications. The data set finally used is genuine, 'real' data and contains variables which approximate closely to those collected in GB in 1981 : therefore it is a near-ideal solution to the problem. These data were made available to us as a result of previous work in classifying individual - level census data carried out by one of the authors (SO) and colleagues at Newcastle University for a census agency in another European country.

A data base pertaining to 481,165 households and 1,228,068 non-UK individuals has been set up on disc. Details in the population file include unique id., census (ED) area code, age, sex, marital status and occupation of the individual concerned. For the household file, the details stored include unique id., census area code, locality, occupancy, tenure, number of rooms, and presence or absence of toilet, flushing toilet and bathroom. Since the great bulk of accesses to census data are on a within - county basis (indeed the data are distributed by OPCS on a county by county basis), it is logical to partition the data set on this basis and the files created approximate to the characteristics of a 'typical' English county. Thus the test data set provides a good approximation in access times, though not in total data volumes, to a national data base.

AVOIDING DISCLOSURE OF CONFIDENTIAL INFORMATION

Central to any success that our system might have is the ability to guarantee that no information will be disclosed on identifiable individuals. Failure to ensure this would render impossible the use of the system by users outside of OPCS, at least directly over telecommunication lines. In addition to a need for the technical demonstration of security, the public perception of such access to confidential data must also be satisfactory. This paper is, of course, mainly concerned with the technical problems but plans were made for independent testing of the system (see objective (iv)).

The problem of non-disclosure has been treated both theoretically and empirically (see, for instance, Duncan and Lambert 1986, forthcoming; Bethlehem, Keller and Pannekoek 1988; Redfern 1988). So far as this project is concerned, our objective is to release no aggregate data which are such that a single individual can be identified and characterised, even by repeated querying of the data base and subtraction of the results of slightly overlapping queries. We took a policy decision that the prevention of disclosure could not be based upon comparison of requests with all previous requests (though maintaining a log of all requests is part of our system); thus, all statistics must be produced in such a fashion that security is preserved irrespective of any comparison made with previous output. More positively expressed, we seek to do the minimum to the data which will retain confidentiality, thereby maximising the value of the data. Strictly, the problem of identification of any one individual from a micro-data set (such as the US Public Use Sample) is not our concern unless this system is used for generating (perhaps purpose-specific) microdata sets.

Four obvious techniques exist for ensuring that no disclosure occurs in aggregate data of the kind hitherto released in Britain :

- deletion of data where the value(s) of some variable(s) do not exceed a defined threshold

- banding of the data into groupings (e.g. with counts ending only in 0s or 5s)

- the introduction of 'noise' by adding small random numbers and the publicising of this element of randomness

- adjusting the area of aggregation in response to the frequency distribution(s) of interest.

Of these, the first and third have been used routinely in the UK in recent censuses and, in the manner implemented, have been accepted without significant debate. In the Small Area Statistics, for instance, it has broadly been the practice to provide missing data values for all bar the most basic few cell counts unless the resident population was 25 or more people (population records) in the area concerned or unless there were 8 or more households (in the household records) therein. All cell counts other than those few remaining in suppressed records were then subject to adjustment through the addition of +1, 0 or -1 (Denham and Rhind 1983, p. 81). The amalgamation of adjacent areas to produce statistically sound and confidential aggregates has also been followed by OPCS in dealing with the 10% sample data and with the 1971/81 Change Files.

Precedent is a valuable matter in dealing with confidentiality : moreover, we know of no circumstances where this has been breached in previous censuses. Hence, our initial experiments have used data deletion and addition of 'noise' and have adopted, initially at least, the same thresholds as OPCS' previous practice. The effect of the application of such rules is, of course, highly data-dependent. Of the 125,000 Enumeration Districts in the the

1971 SAS, the data for only 800 suffered suppression whilst of the corresponding SAS data for the 149,000 populated 1km grid squares, no less than 54% contained suppressed fields (though these related to less than 5% of the total population). This difference reflects the difference between the low variance in population in areal units chosen to have approximately equal numbers to even out the enumerator's workload and the massively skewed distribution in geometrically regular areas. Clearly the size and shape of the areas requested, as well as the variables chosen, have major effects upon the accuracy and utility of the resulting data if confidentiality is not to be breached.

THE INITIAL, HARDWARE-BASED SOLUTION

Introduction

Our initial intention was to create a simple and easily changed prototype which would merely demonstrate the technical feasibility of the concepts. We intended to build this using a Fourth Generation Language for ease of modification: performance was not an important criterion at this stage. This was to have been carried out on university facilities and, in particular, on ICL equipment both to exploit the characteristics of the Content Addressable File Store (CAFS) and to replicate at least part of the long - standing OPCS/GRO(S) computing environment.

In the event, delays in the ESRC funding serendipitously proved helpful because ICL announced INDEPOL, a highly secure software system initially created for handling individual records for the defence sector and for the police. The advantages of INDEPOL over other, less hardware-specific solutions stem substantially from its exploitation of CAFS. In the first instance, the latter provides:

- fast searching at the disk. In principle, this permits searching at up to three megabytes per second on each disk and places little load on the central processor.

- multiple criteria searching: up to fourteen search criteria may be used simultaneously, with boolean or quorum logic using precise, stem, or fuzzy matching.

- the highest level of computer security yet available in the public domain. Access control by user, by terminal, by data name and data value, by use of passwords and other methods is claimed to ensure that, for instance, a particular 'high clearance' user can be prevented from performing functions which are legal in his own office whilst he or she is using an insecure terminal. The user cannot find out anything about data or facilities to which s/he does not have access: their very existence remains hidden.

- the ability to create menus and templates but also to make a query using a free format enquiry language. Macros may be shared and made generally available for frequent enquiries.

- ways for rules to be embedded in the system.

The reason for our selection of INDEPOL is not merely all of the above reasons, important as they are. In addition, it facilitates the move from the need for complete pre-planning of software to a situation where all the significant details of an application are held in the System Model and the Data Model: these can readily be edited - even within a single log-on - and, since the models are used interpretatively, there is minimum dependence on pre-compiled code.

The existence of INDEPOL - albeit at an early state of its product development cycle - therefore enabled us to revise our strategy. It enabled us to concentrate on building a Census - specific superstructure and, at the same time, to take advantage of the highly secure hardware and software facilities, including networked access, devised by ICL for their original clients. All of the work using INDEPOL was carried out in ICL's Defence Systems Division in Winnersh, Berkshire, UK.

Results

Our experience with the INDEPOL system is that it works extremely well for standard enquiries and simple presentation of either listed or singular results. The feasibility study has shown that individual queries are speedily answered, that tabular reports can be constructed and printed out on a peripheral device and that record- specific constraints can be applied to suppress or amend answers to queries. Moreover, standard facilities provide bit-compression of data fields wherever possible, minimising the data storage required. In essence, only two of the requirements of the system were not demonstrated satisfactorily : the addition of secondary geographical indexes (which, from seeing other INDEPOL applications, we are convinced would work well provided no more than about 10 indexes were required at any one time) and the public demonstration of the level of security (see below). More specific conclusions from the study include :

(i) INDEPOL has a straightforward command language for querying the database; even complex cyclic control programming can be quickly mastered by the application designer.

(ii) It is extremely secure and very versatile in the way in which applications, system and user security can be manipulated in terms of restrictions by data name, data value, command availability and option availability. In addition to controlling access, the system keeps complete records of all requests and commands. However, the sole availability of INDEPOL in a highly secure environment at the time of our tests, together with the need to work on Model 204, ensured that we could not carry out our initial intention - to throw open the system over the academic network with a challenge to colleagues to break into it!

(iii) The CAFS system makes the INDEPOL database extremely fast (as fast as the disk control unit). Typically, a standard enquiry searching 1.2 million records and cross-linking two files took rather less than 30 seconds. Many inquiries on a single file took only 2 to 4 seconds. A loop structure using cyclic control to construct tabular reports took less than 90 seconds to search 0.4 million records a total of 55 times and print the results.

(iv) The design of the INDEPOL data model is straightforward in terms of the relationships between entities and attributes. The actual coding of the data model does require a certain amount of data processing expertise but can be quickly mastered by an experienced applications designer. From the user's point of view, how attributes relate to one another and to the physical world is evident from a brief study of the data model.

(v) Template and menu design is very simple and there is an extensive validation ability which is always present in template construction. This allows for complex applications to be built up very quickly by even the most inexperienced of users. Use of the templates also permits querying of the database with only a limited knowledge of the INDEPOL command language.

(vi) The ease of adding attributes _post-hoc_ to the data model means that secondary features can readily be added to the system at a later date (e.g. the addition of either post-code or Ordnance Survey grid references to provide locational references) with the minimum detrimental effect on the performance of the system.

The majority of the problems encountered with INDEPOL stem directly from our use of the software for tasks for which it was not optimised or designed. The main problem areas are:

(i) The production of two-dimensional tabular reports presents the major problem. On-screen production of tables is limited to 23 lines of text with no ability available to page through a large table (though it is possible to print the table through a spool file).

(ii) In submitting tabular reports to the spool file, a large number of SCL and COBOL programs plus relatively complex and convoluted INDEPOL macros are required. An obvious means of reducing these problems would be to re-design the SAS tables into a series of standard layouts or choose a database system that combines the query ability of INDEPOL with a bolt-on report writing facility similar to RPT/RPF in ORACLE.

(iii) There are only very rudimentary algebraic, and practically no statistical, manipulation (including sampling) capabilities in INDEPOL other than the use of standard operators and functions such as SUM and MEAN. It is envisaged that a 1989 release of INDEPOL will go some way towards catering for these particular requirements.

THE SOFTWARE - BASED SOLUTION

Work involving Model 204 is, at the time of writing, at a much earlier stage of development than that of INDEPOL. Initial familiarisation with the MVS/XA operating system running on the Amdahl 5890-300E computer and with the Model 204 software (which is currently used only by the Librarians in Manchester), together with the need to re-load and index the test data, has ensured that initial demonstrations of a primitive system could not be given until end-January 1989.

Nonetheless, some preliminary conclusions can be made though it must be appreciated that these may reflect inexperience with Model 204's capabilities. These are as follows :

(i) Model 204 is highly unlikely to provide as secure a system as INDEPOL or any system exploiting CAFS.

(ii) Similarly, it is unlikely to provide as rapid response to simple queries as INDEPOL unless much indexing has been carried out; such indexing appears extremely expensive on storage space. INDEPOL scores heavily in situations where unanticipated queries are made.

(iii) Model 204 is however relatively easy to use, has SQL and has the benefit of good report generation capabilities. For production of complex standard tables, it may well be more efficient in terms of both human and computer resources than INDEPOL.

CONCLUSIONS

All of our work is predicated upon the notion that the data base is run within the Census offices and accessed by one means or another. Given this, the relative merits of the two solutions seem likely to be exploitable in different modes of operation. Thus it seems likely that technical considerations of an on-line querying system could be satisfied by INDEPOL; this of course does not address directly the question of public perception of the risk of disclosure. Based upon our experience to date, the following scenarios might be feasible ones :

(i) Use of INDEPOL (or, at least, CAFS-based systems - CAFS units are now available on SUN systems) to provide an on-line enquiry service over national telecomms networks. Conceptually, this would be simply an extension of the highly successful National On-line Manpower Information System (Nelson and Blakemore 1986). It would be viewed as a parallel facility to the distribution of conventional statistics for standard areas (see below).

To make this work, groups of users would have to be authorised in the same way in which over 200 existing NOMIS groups have been authenticated. These include central and local government, the commercial sector agencies and academics. All who do not have 'as of right' access (as local government does to some datasets held on NOMIS) would require licensing.

Users would therefore access the 'raw' questionnaire responses but would only be permitted to receive aggregate output, suitably screened to ensure non-disclosure. We are quite confident that this could be achieved by connecting the CAFS machine directly to the network though constraints exist in terms of the type of terminal needed if INDEPOL is used. Turn-around of the order of seconds or, at worst, minutes would be available by this approach.

(ii) Use of Model 204 to run a service based upon OPCS receiving queries in one of a number of forms; the best solution would be for users to compile queries and dispatch them over telecomms to OPCS who then transfer these to a non-networked machine on which the database resides and produce the results required. Depending on the mode of transfer, this will provide responses in the time scale of an hour to several days.

Both of these scenarios would require the appointment of a data certification officer who would be responsible for monitoring the flow of requests. Both would benefit from the construction and low-cost distribution of a Query Formulator for PCs which would enable users to design their requirements in a way suitable for the system and for transmission to OPCS, with obvious syntax and logic errors being weeded out at source. Equally, both could have advantages for OPCS and GRO(S) not yet mentioned : they would diminish the penalties for failing to anticipate user demand correctly, they might facilitate the construction of new products such as Public Use Samples and they would simplify the design of samples for other surveys and data linkage for files held within the remit of the Registrar Generals (e.g. demographic, housing and mortality and morbidity data).

Nothing of what is proposed is totally novel. Credit rating agencies (e.g. TRW in the USA) already operate larger data bases and provide country - wide access to the data. An OPCS - fostered scheme, the Longitudinal Study, already gives a form of on-line access in the UK to anonymised records of sampled individuals. Moreover, this particular proposal has to be seen in the context of many other developments in data collection in Britain, notably the compilation of the Community Charge register by local governments (Redfern 1988). What is critical, however, is the paramount need to ensure that no disclosure of information about identifiable individuals occurs.

The costs of setting up and running such a system are not yet quantifiable in any detail, in part because of reviews of charging mechanisms and the likely introduction of new facilities in the census offices. Assessment of costs and benefits will also depend upon decisions of government on financial recovery targets from production of the census data; as one example, the lowest effort and cost method of distributing standard statistics would be to produce one CD-ROM covering the whole country and sell this to every customer, even if they only required data for one small area (as is generally the case).

By its very nature as a previously unavailable service, the level of customer interest is unquantifiable in any detail. However, in a survey of the academic community (Marsh et al, 1988) no less than 76% of respondents said they would use such a tool (subject to its cost) and a further 22% said they were undecided. The NOMIS experience suggests that, with appropriate training and publicity, a wide variety of users can be expected; for many of them, doing without NOMIS would now seem inconceivable. Finally, availability and success of such a system could have implications for the commercial census agencies licensed by OPCS after 1981; as well as continuing as purveyors of standard statistics to naive or very infrequent users, they would have to become census skills centres and act as intermediaries for any customers who required help in formulating queries to the proposed system.

ACKNOWLEDGEMENTS

The developments described in this report would not have been possible without the help and advice of a number of other individuals. Thanks must first go to Martin Higgins, now of Pinpoint Ltd, for the initial development work. In addition, Mark Musto and Geoff Sprott (part of the ICL INDEPOL Support Team), Dick Morgan (ICL INDEPOL Applications Manager), and also Miriam Glyde and Dick Goodwin (of ICL Winnersh) are owed heartfelt thanks for their support and advice in the earlier stages of development. Colin Wymer kindly made a copy of the data used in the study. In OPCS and GRO(S), Chris Denham, Basil Mahon, David Pearce, Terry Russell and Frank Thomas have been invariably helpful in all matters related to this project. Keith Cole and colleagues in the University of Manchester Regional Computer Centre have been immensely helpful in regard to our use of Model 204.

REFERENCES

Bethlehem J.G., W.J. Keller and J. Pannekoek (1988) Disclosure control of microdata. Proc.4th US Annual Research Conference of Statisticians, 181-92.

Carmichael, J.W.S. (1987) INDEPOL : A software package exploiting CAFS. ICL Defence Technology Centre, Winnersh

Denham C.J. and Rhind D.W. (1983) The 1981 Census and its results. in Rhind (1983), 17-88.

Dewdney J.C. and Rhind D.W. (1986) The British and USA's Censuses of Population. in M. Pacione (ed.) Population Geography : Progress and Prospect, 35-57, Croom Helm, London

DoE (1987) Handling Geographic Information. The report of the Committee of Enquiry chaired by Lord Chorley. HMSO, London.

Duncan G.T. and Lambert D. (1986) Disclosure-limited data dissemination. J. Amer. Statist. Assocn. 81, 393, 10-28.

Duncan G.T. and Lambert D. (forthcoming) The risk of disclosure for Microdata. J. of Business and Economic Statistics

HMSO (1988) 1991 Census of Population, Command Paper 430

Marsh, C., Arber S., Wrigley N., Rhind D. and Bulmer M.(1988) The View of Academic Social Scientists on the 1991 Census. Environment and Planning A, 20, 851-89.

Nelson R. and Blakemore M.J. (1986) NOMIS - a national GIS for the management and mapping of employment, unemployment and population data. Tech. Papers 1986 ACSM/ASPRS Annual Convention, Vol. 1, 20-30.

OPCS (1988) 1991 Local Statistics : a discussion paper, Office of Population Censuses and Surveys, London.

Redfern P. (1987) A study on the future of the census of population: alternative approaches. Eurostat Theme 3, Ser. C. Luxembourg : Office of Official Publications of the European Community.

Redfern P. (1988) Population registers : some administrative and statistical pros and cons. J. R. Statist. Soc. A 151, 3.

Rhind D.W. (1983) ed. A Census User's Handbook, Methuen, London.

Rhind D.W. (1984) The SASPAC story. BURISA, 60, 8.

Rhind D.W. (1985) Successors to the Population Census. Jl. Economic and Social Measurement, 13, 29-42.

Wise S. and Pellatt R. (1988) Rapid access to a large cartographic data base using CAFS : a feasibility study. Tech. Rept. 4 in GIS, Computing and Cartography, Univ. Bath