
Hands on the Census: Microdata from the 1991 Census of Population in Britain

by Catherine Marsh¹
Department of Sociology and
Department of Econometrics
University of Manchester

The information dilemma

A basic dilemma confronts the producers of data in the public sector. On the one hand, they face users demanding that information collected at public expense be made available to the research community in ever increasing detail. On the other hand they face those concerned with privacy and confidentiality worrying about the possible release of identifiable information about individuals. In short, they have to reconcile competing moral claims: they are caught in the middle between one group of citizens asserting a right to information and another group asserting a right to privacy.

The audience at this IASSIST meeting, as professionals in the business of information acquisition and dissemination, will have rehearsed the arguments many times about the rights to information. It is important to remember that the arguments for a right to privacy and confidentiality are also strong, but have changed their character somewhat during the last decade, and have had a profound influence on the preparedness of state authorities and census agencies to release microdata to the research community.

The state is no longer the only, and perhaps no longer even the major, collector of systematic personal data on individuals. The expansion of information technology has led to a new private industry of information collection and management. Many companies take as their base the publicly available state collected information such as electoral registration lists. Some then may link to this other published records e.g. on bankruptcies and criminal records. Others specialise in the collation of information from pull-out questionnaires in magazines and so-called guarantee registration cards, others bring together personal financial information from major credit card companies and chain stores. Sometimes the aim of this data collection is to assess the credit-worthiness of individuals. Sometimes it is to facilitate modern direct marketing techniques, targeting prime areas for particular mail-shots. Combined with advances in telephone technology, these techniques have become big business. In Britain, several companies have datafiles containing information on over 90% of households. Most of this information is referenced by a combination of name, address and post-code.

There is widespread public concern about the activities of these commercial companies. The British public is very hostile, for example, to the sale of election registration lists to outsiders (Campbell 1987). In the lead up to the 1991 census, there were media programmes and newspaper articles which expressed worry about the potential of census information about small areas to be linked to other private information databases both of these private companies. When the recent Census (Confidentiality) Bill was being debated in Parliament, certain members suggested amendments which would have made the linkage of census information to fine-grain post-coded information illegal (Computer Weekly 31 Jan 1991). The attempts did not succeed, although plans to release statistics for post-coded areas in England and Wales were modified.

Both the political right and left have turned to the state to protect citizen liberties and rights to privacy. Throughout the 1980s, most European countries, unlike the US and Canada, have enacted legislation to give citizens rights with respect to databanks of information which may be held about them. The Data Protection Act of 1984 gives individuals in the UK rights to find out what computerised personal information any organisation might hold about them, to challenge it if wrong, and claim compensation if they suffer as a result. Owners of machine-readable lists are obliged to register any file of identifiable personal information they hold, and to say for what purposes it is held; the register is open for public inspection.

The Data Protection Registrar attempts to police the activities of the information industry; he has, for example, recently attempted to curb the unrestricted use of address-based information for credit-redlining. However, in some ways, the individualistic focus of the existing data protection legislation weakens it as a tool for those seeking to ensure that information released could not be linked to private databanks. Rights to know are restricted to rights granted to individuals to find out what is held about *themselves*; there is no right for someone wanting to establish how much census information it is safe to release to obtain answers to such critical questions as how many individuals are covered in the databank and which census variables are held which

would be available for matching purposes. The format of the Data Protection register does not enable it to be used as a source for answering such questions. Nonetheless, in the Office of the Data Protection Registrar there exists a team of individuals whose job it is to know precisely who has what computerised information about whom, and to police the workings of the Act; in practice they know in broad outline the information gathering activities of all the major data collection companies.

As well as worry that information may be passed on to the commercial sector, the other major public concern about census and survey data relates the use to which the government itself might put the information. Worries have been expressed in Britain at the time of previous censuses about the passing of identifiable census information to immigration officials or to social security officers. A similar set of concerns were expressed in 1991 over the possibility that census information might be passed to Community Charge officers, responsible for compiling lists of the adult population in order to collect a new and very unpopular flat-rate local tax based on a head count (the "poll tax"). Some campaigners against the tax (e.g. in Tottenham in London) explicitly called for a boycott of the census on these grounds. (It is important to emphasise that the census authorities would never in fact pass on census data to any outside the census office, even to other government departments.)

Taking a comparative perspective, it does appear to be the case that the British public is more sensitive to confidentiality issues and less willing to trust the census authorities than in other countries. The British public is more concerned about privacy than other Anglophone countries; Goyder and Leiper (1985), for example, did a content analysis of letters to the press about the censuses

in 1980 and 1981, and found much more concern over issues of privacy and confidentiality in Britain than in US or in Canada.

Given the concern about confidentiality, it is worrying to learn that disbelief in the absolute confidentiality guarantees given by census authorities is widespread. This is illustrated by a recent Gallup survey undertaken in both GB and USA. It documents that the level of trust in the census authorities is low in both countries, but lower in Britain: the survey shows that the British public is much less likely to believe the confidentiality pledges given by the census offices than the American public, as Table 1 shows.

In short, the general public has a range of worries that census information, gathered ostensibly for assisting government in planning purposes of various kinds, will be circulated to others for purposes which were not declared at the time when the information was collected. In lights of these worries, it is not surprising that census offices do not just hand out microdata on request. It is also not surprising that the census offices that made the decision to release microdata earlier on were more liberal about what they were prepared to release than those trying to make the same decisions more recently.

The resolution of the dilemma with respect to census data

Faced with the dilemma between rights to information on the one hand and rights to privacy on the other, the census bureaus in different countries have made different responses. In general, the English-speaking countries have tended to give primacy to rights to information, and have made census microdata available in various anonymised and sanitised forms, whereas European

Table 1: Perception of census confidentiality in Britain and US

Question: "How confident are you that the Census Office will not release an individual's census information to other government agencies: are you ...

	GB	USA
	%	%
very confident	17	23
somewhat confident	27	44
or not at all confident?"	42	28
(don't know)	14	5

Fieldwork dates: March 1991 for GB; March 1990 for US;

Source: Gallup Political and Economic Index, No. 368, April 1991 (GB) and Gallup Report for USA

Reproduced with the kind permission of UK Gallup.

countries have been much more exercised about rights to privacy and have generally not gone down this public road. Britain, standing as it does with a foot in either camp, has taken a long time to decide which way to go.

The United States of America was the first country to release such public use files in the 1960s. From the time when the population census was first computerised in the US, discussion was initiated about releasing forms of microdata to the research community. Partly because the administrative culture was open to research dissemination, and partly because of the existence of energetic individuals pushing from within the census bureau, "public use files", as they are termed in the US, were released retrospectively for the 1961 census and have become a routine part of census output. They were followed by Canada in the 1970s and Australia in the 1980s. However, Canada and Australia never released as much information, either in terms of sample size, detail of file structure or fine-grain detail of coding schemes as the US public use files (see Marsh *et al.* 1991a for more details). The Australian census microdata in particular contains only the state as a geographic identifier.

Despite similar requests for microdata, many European countries have declined to release samples into the public domain unfettered by limitations placed on uses or users. Academic researchers in Denmark and Sweden may only receive microdata for specified and delimited purposes. Some countries allow local state authorities access to census data but deny similar access to academics: local government researchers at all levels in Italy can have access to microdata, for example, as can regional governments in Spain and some government departments in Luxembourg. In Germany, anonymised census records may only be released to the communities². And some countries only release a small subset of the available information in the form of microdata; in France, for example, only a restricted subset of census variables is released publicly as microdata in sampling fractions varying from 0.1% to 25%. (Redfern 1987 briefly outlines the rules in each European country.)

I am delighted to tell you that for the 1991 census in Britain, agreement has been reached for samples of anonymised records to be released. At present the plans only include England, Wales and Scotland, but representations are also being made to the Census Office in Northern Ireland to grant a similar request with respect to the 1991 Northern Irish census.

Requests from British academics for census microdata go back at least fifteen years. Spurred by interest in the release of public use samples in North America, a committee of interested academics was convened in the mid-1970s to discuss the possibility of obtaining similar

microdata in Britain for the 1971 census. One problem which emerged early on was the geographers' demand for microdata was for very large samples (10% or more of census records) with very fine grain geography. During the 1970s, it seems that the needs of geographers dominated the requests, and the gulf between what academics seemed to require and what the census offices felt able to release while retaining confidence in the confidentiality of the records was wide.

Furthermore, the early discussions about microdata never progressed very far since the legality of releasing microdata under the terms of the 1920 Census Act was never resolved. There were those inside OPCS who argued that release of microdata was permissible under the terms of the Act (Redfern 1976), but the legality of this move was contested by others at the time.

More concerted efforts were made to obtain microdata from the 1981 census. The White Paper outlining plans for the 1981 census made it clear that the census authorities were prepared to consider reasonable requests for samples of anonymised records. Support for release of microdata was also available from other sources, some of them somewhat unexpected: from the British Computer Society team reviewing security provisions for the 1981 census (HMG 1981a), and from these advocating cuts in the Government Statistical Service who argued that microdata could provide a cheap and flexible substitute for tables (HMG 1981b). There was also interest was shown by census office research staff (Denham 1986). And requests from academics (e.g. Norris 1983) for microdata persisted.

A further committee was therefore convened under the auspices of the Environment and Planning Committee of the Economic and Social Research Council between 1984 and 1985, to try to co-ordinate a request to be put to the census offices to obtain retrospective samples from the 1981 census. Despite receiving evidence from several academics about the value of such data, the committee never reached the stage of putting a formal proposal to the census offices. There appear to have been several reasons for this. First, the demands of the geographers who wanted fine grain areal information to the detriment of detail elsewhere and the demands of social and policy researchers who wanted full household information, if necessary sacrificing geographical detail were never reconciled. Second, it proved very hard to get agreement from others in the commercial and public sectors to form a purchasing consortium to buy the proposed data. Third, in 1985 it seemed likely that there might be a major 10% household survey undertaken in 1986 which might meet needs more effectively; (in fact this survey was never undertaken). As time furthermore, as time went by, the value of data relating to 1981

seemed to decline³. The request for data from 1981 therefore lapsed.

However, user demand did not (Marsh *et al.* 1988). The Economic and Social Research Council therefore renewed its efforts in good time for the 1991 census, and set up a working party to negotiate with the census offices in Great Britain and to present a formal request. This working party undertook some systematic work on quantifying the risks of disclosure from releasing census data, and concluded that the risks were minimal. On the basis of that work, it proposed that the census offices release two different files of microdata, one to meet the needs of those who wanted the maximum geographical detail and one to service those whose prime interests were in household structure. The request was presented in a lengthy report in 1989 (published in Marsh *et al.* 1991) which was favourably received by the census offices. The committee also secured the agreement of the ESRC to shoulder the total costs of the purchase if necessary.

The census offices sought advice from their solicitors about whether microdata could legally be released under the terms of the 1920 Census Act. They were advised that anonymised microdata came under the general heading of a 'statistical abstract', and could be released without changing the law. The Office of the Data Protection Registrar and Liberty (the National Council for Civil Liberties as was) were consulted, and neither had any major objections. The proposal to release microdata was therefore mentioned in the White Paper outlining plans for the 1991 census (Cm 430, 1988), subject to the overriding need to preserve census confidentiality. It was also commented upon by the British Computer Society team who reviewed security arrangements for the 1991 census (Her Majesty's Government 1991); since they had supported the idea for the 1981 census, they gave the idea their blessing. In July 1990, the agreement in principle of the Registrars General to the ESRC request was announced in a written Parliamentary answer.

Background to the British Census

In the United Kingdom, censuses of population are the responsibility of three separate offices: the Office of Population Censuses and Surveys under the supervision of the Registrar General for England and Wales, the General Register Office for Scotland under the supervision of the Registrar General for Scotland, and the Census Office in the Northern Irish Department of Health and Social Services under the supervision of the Registrar General for Northern Ireland. The Office of Population Censuses and Surveys (OPCS) plays a coordinating role in the work of the three offices.

The legal framework for censuses in Great Britain is the 1920 Census Act as amended by the Census (Confidentiality) Act 1991. The 1920 Act makes provision for population censuses to be taken at no greater frequency than every five years. It is enabling legislation which requires there to be a new census order outlining arrangements and plans for content each time. The Registrar-General is given the responsibility for organising the census, the power to take on the necessary staff, and the right to present the final accounts to Parliament.

Under the terms of the 1920 Census Act, filling a census return is compulsory for householders. Many in census offices feel that the compulsory nature of the census puts it in a different moral category when it comes to the release of microdata. The logic of this position is not entirely clear, however, since similar confidentiality guarantees are given to those who take part in voluntary government surveys⁴.

The Registrar General is given the duty to ensure that summary reports of census data re prepared.Furthermore,

"The Registrar-General may, if he so thinks fit, at the request and cost of any local authority or person, cause abstracts to be prepared containing any such statistical information, being information which is not contained in the reports made by him under this section and which, in his opinion it is reasonable for that authority or person to require, as can be derived from the census returns." [Section 4(2)].

The interpretation of this section of the Act was critical for the release of microdata; it turned on whether microdata could be deemed a "statistical abstract".

Both by international standards and by comparison with previous British censuses, the censuses of 1981 and 1991 were fairly slim. In 1991 there were 8 questions about housing and 19 questions about individuals. There has never been an income question nor, since 1851, any question on religion in Great Britain, although there is such a (voluntary) question in Northern Ireland. Enumeration is done on the basis of presence on census night; information is also obtained about the usual place of residence of visitors and about absent usual residents (with a voluntary return if the whole household is absent).

There are no long and short forms on the British census. Sampling is only undertaken at the processing stage. The answers to those questions which are laborious to code (such as occupation, industry and qualifications) are only fully coded for a 10% sample.

When it comes to census output, the principle of dissemi-

nation is best expressed in the motto of the London Statistical Society, first expressed over 150 years ago: *alitis exuerendum*⁵. At the time it was first enunciated, it was an unattainable ideal, as the technology for taking censuses and surveys had not progressed very far, and the information was usually published in the form of verbal commentaries, often of a very opinionated form (Cullen 1975). However, the trend in census output has continuously been towards making more and more of the detailed information available, aiming for the user to be free to rework it and interpret it in any way that he or she chooses. First this was achieved by providing more tables. There will be 20 volumes of special topic statistics from the 1991 census and monitors for each county and parliamentary constituency. More recently, increased volume of information have been supplied in machine-readable form; the census offices estimate that 98% of census output nowadays is in electronic form. In particular, the small area statistics are particularly detailed; information is provided down to aggregates of around 200 households in England and Wales (70 households in Scotland) and around 9,000 pieces of information are available at this level. There are also special workplace and migration statistics released in machine-readable format for small areas.

Thus, up to 1991, the output from the British census in forms other than microdata was exceptionally detailed, especially in the amount of machine-readable data made available for small areas. This may also be part of the explanation why the demand for microdata in the past never became overwhelming.

The information to be released

At the time of writing, agreement in principle has been given to a plan to release two different samples of data from the 1991 census. Their broad structure has been agreed, but negotiations are still continuing about the precise details. Any of the details mentioned here could therefore change before release of the data.

The first file proposed is a two percent sample of individuals with full housing information and some limited information about household structure attached. The census offices have been guarded about releasing too much information about other household members for fear of effectively releasing what amounts to a hierarchical file under a different guise. This file will show a geographical scheme identifying large local authority districts or groupings of smaller ones.

The second file proposed is a one percent hierarchical file of households, containing housing information and full information about all the individual household members. For confidentiality reasons this file will classify data only to the 10 Standard Regions of Great

Britain.

Both files are to be drawn from the 10% of records which are fully coded. The sample will be drawn systematically from this 10% file which is ordered by county, then by enumeration district and street. The two samples will be drawn without replacement so that there will be no overlapping subsamples.

Measures to protect confidentiality of the information
Great effort is put into ensuring that the data is safe from identification and disclosure. Five different devices are being used.

(i) Sampling of records

Sampling itself is one of the most effective ways of reducing the risk of disclosure, provided that users cannot identify which individuals are selected. The sampling fractions are small (.01 and .02), and, because the geographical identifiers are different on the two files, it will not be possible to combine them.

(ii) Suppression of variables

Names and addresses are not entered onto the census computer, and therefore obviously not even available for suppression. Precise data of birth is to be suppressed; age will only be available in yearly bands, and will be top-coded (see below). There will also be no information on the imputed missing household⁶, since these are excluded from the 10% coding operation.

(iii) Limiting geographical detail

There were competing claims about the basis of SAR geography. The main choices were:

- local government administrative geography
- health service administrative geography
- political boundaries
- local labour market areas
- postcode sector defined boundaries
- grid square geography

It was not possible to have more than one system, as the small overlaps between many of these different schemes would have led to unacceptably small areas being identified. On general utilitarian grounds of maximum benefit to maximum numbers, local government administrative geography was chosen.

The geographical scheme in the individual level file will identify only those local districts with populations of

120,000 or more; the rest will be grouped into areas of at least 120,000 population. However, all but one of the metropolitan local authority districts will be identified. On the hierarchical file, the geography will be even coarser. Only standard regions will be identified, with London subdivided into inner and outer areas; this amounts to twelve areas in all, the smallest of which East Anglia with an estimated population of 1.9 million.

Other geographical information is also collected on the census - the usual address of visitors to the household, workplace address, students' term-time addresses and address one year ago. These are to be very heavily restricted; at the time of writing, the proposal is to identify standard region, a same district/different district identifier and perhaps a distance measure as well.

The order of records in the microdata file are to be scrambled to ensure that locality information cannot be obtained from the proximity of one records to another.

(iv) Grouping of categories

A rule is being used to restrict the fineness of the coding scheme for each variable:

Each category of each variable to be identified on the SAR must have an expected sample count of at least 1 in the smallest geographical area permitted on either file.

Expectations are formed by scaling down the distribution of that variable for Great Britain as a whole. To illustrate, in the individual file, the smallest geographical area identified will have a population of 120,000. The sampling fraction will be 1/50. Thus the cut-off on this file is 50/120,000 times the population of Great Britain of 56 million, yielding 23,3000; this is then rounded up to 25,000. Accordingly, any category of any variable which in Great Britain is expected to have less than 25,000 people at the census will be grouped in with another category in the SAR. In the household file, the smallest region identified is East Anglia, with a population of around 1.9 million people, and the sampling fraction will be 1/100, so the cut-off being used for this file is 2,700 people.

Thus only the univariate distributions are used to identify categories at risk. Some have been worried that it is unusual *combinations* of categories that cause problems: female barristers, small householders in accommodation with a lot of rooms, and so on. However, some recent work suggests that the strategy of worrying only about those categories that are small in the univariate distribution will predict people who have unique combinations of variables extremely efficiently (Marsh *et al.* 1991b).

The variables most affected by the grouping rule are occupation (where the 350 categories in the full coding scheme will probably be reduced to around 250), industry (270 categories reduced to around 200), detailed educational qualifications (103 reduced to 53) and country of birth (reduced to 47 groups). The variables to be top-coded are age (to be grouped into two year bands between 91 and 94 and top-coded 95 and above), hours worked last week (top-coded after 70), and number of rooms in the accommodation (to be top-coded after 14).

Top-coding does not solve the problems of households with very large numbers of individuals; the number of people in a household affects the very structure of a hierarchical file rather than the categories of one variable in it. Several suggestions are currently being explored to solve this problem, including removing all geographic identifiers from households with more than 12 members.

(v) *Perurbation*

The small area statistics from the British census have always been subject to a degree of random perturbation ('Barnardisation') whereby a random +1, 0 or -1 is added all cells. The comprehensive application of this technique has always been unpopular with users (Marsh *et al.* 1988) and the cumulative effects of the small errors introduced can pose quite serious analytic problems (Senior and Cole 1991). Wholesale Barnardisation is of course not possible with microdata, and the prospect of deliberately adding more noise to many different variables on top of the natural levels of error already existing in the data was resisted by ESRC team negotiating the release of the data. Instead, the census offices have suggested that a small number of individuals in each area be switched with others in a nearby area (Griffin *et al.* 1989), a technique which amounts to adding a small degree of random noise to the geographical identifier, but preserving the rest of the household data intact.

In the USA and Canada, the census offices appoint some sort of panel to oversee the arrangements for protecting confidentiality (Gates 1988). Such a Microdata Review Panel was considered for GB, but, on the recommendation of the Royal Statistical Society, the census offices have appointed one technical adviser instead, to oversee the specification of the SAR files and the general arrangements made for confidentiality. This advisor, a senior academic statistician, will make independent recommendations to the Government Minister responsible for the census offices.

Contractual arrangements

The funders of the project are the Computer Board⁷, <footnote text> who are paying for the data, and the Economic and Social Research Council, who are putting up the money for a research and distribution centre to be

established to house the data.

The contract to buy the data from the census offices is regulated in part by legislation; under the terms of the 1920 Census Act, the census offices are obliged to recoup the cost of producing extra tables (which includes the SARs), but are prohibited from making profits on the data which they supply. The ESRC team negotiating the release of SARs attempted to find co-purchasers to enter into a consortium to share the costs, but none came forward. Thus now, the academic purchasers are bearing the full developmental costs of the SARs. The census offices have done their costings on the basis of passing on the full marginal costs of producing SARs; at present the cost of the data seems likely to be around £200,000, but this sum will only be finalised when the file specifications have finally been approved.

The contract to must safeguard the interests of the academic purchasers in their product. Although final contractual details are still being sorted out, we hope that, in return for our payment, we will get full exploitation rights of the dataset, which will be sole rights for a limited period. Since the academic purchasers of the data are bearing the full cost of the production of the files, the census offices will not receive any further royalties when value-added products of the census are passed on to third parties.

The contract to purchase from the census offices will specify that all end-users of this data must give various undertakings about respecting census confidentiality. People will be expressly forbidden to try to identify individuals in the SAR, link them with other sources of data or to claim to have done so. Any breach of this agreement would lead to withdrawal of the SARs. While not undoing the harm caused by a breach, but its threat would probably act as a deterrent since academics would subsequently be unable to publish any information based on the SARs. Various methods are under discussion to register users and keep track of copies. It is possible that heads of departments will be required to be the data holders rather than individual researchers.

The contractual arrangements between the purchasers of the data and the end users with respect to have yet to be discussed in detail. Academic users wanting the data for research purposes will have free access to the data, but means will be sought to regulate the other ways in which the data may be used, to safeguard the interest of the public funders. A graduated scale of charges seem likely, charging commercial users what the market will bear, perhaps with lesser charges to the rest of the public sector and the voluntary sector. University and college faculty who contract their services as consultants outside the academic sector will be charged for their use of the

data.

Disseminating the data to users

The Economic and Social Research Council is funding a centre at a university location to house, disseminate and act as a research focus for this dataset. Invitations to house this centre were put out to several institutions, and three teams submitted tenders. The decision about which will be successful is expected in late May 1991.

Four types of usage are envisaged:

(i) On-line access over the academic network

The Joint Academic Network (Janet) connects all universities and many polytechnics and other institutions in Britain. It is funded by the Computer Board, and free at the point of use to all universities. One important method of giving access to the data will be to mount versions of the data at central locations on the network for easy access by any academic users.

The decision about which software to use for the data has not yet been made. SPSSX seems the most popular candidate for the individual file. The hierarchical file of households could also be set up in SPSS, but this might prove cumbersome, especially since there are two different hierarchies within the dataset: individuals can be grouped into either households of families. SAS, SIR and Oracle are other candidates which might be considered. It is likely that eventually the same information may be held in different formats at different locations; while some may view this as inefficient, from the user's point of view there are great gains in terms of familiarity and ease of use.

(ii) Tables service

The census is a benchmark source of social data. It provides denominators for many different researchers' numerators. However, it is not the prime data source for more than a few researchers. It is therefore important for there to be a service to academics which would provide a service of this kind to other academics, although the demand for this may lead to the need for rationing.

(iii) Customised subsets

Another central task that the service centre will undertake is to extract subsets of variables and cases for different users. This will need to have regard to the media most likely to be demanded; a PC/workstation platform is the most likely here, although demand for data on CD Rom and other media will need to be monitored carefully. There will doubtless be demand for teaching datasets for use in schools and colleges.

(iv) Passing on whole dataset

For efficiency reasons, other academic users would be

encouraged to use the service provided over the academic network. But the whole philosophy of providing samples of anonymised records is that users will be free to port them into their own hardware and software environment and not be restricted by earlier decisions. No restrictions will therefore be placed on others wanting entire copies of the dataset.

Public and commercial users do not in general have access to the academic network. They also tend to be familiar with somewhat different software to that used in universities; in Britain, for example, the most popular tabulation software used by market researchers is a product marketed by Quantime called Quanvert. One of the things that the academic purchasers of the data may want to explore is licensing a commercial agency to provide an on-line service for the commercial sector.

The centre housing the microdata will have responsibility for documenting the datasets, and for computing the sampling errors and documenting these. It will also need to establish a user group, and disseminate information about the database to users, both in electronic media and by hard copy newsletters.

Conclusion

The existence of microdata from the 1991 Census opens up a valuable new resource to British social researchers. Our social research community has grown used to making do with small area aggregates, and drawing inferences from these; the existence of microdata should lead to some interesting work on ecological fallacies. Survey and market researchers will be in a position to design much more efficient samples to locate specific subgroups of the population. We will have a source of microdata on a badly neglected part of the population, namely those who do not live in private households. Research into different means of classifying families and households should blossom given the richness of the hierarchical information available.

Perhaps also we may hope that effort will be made to construct internationally comparable files of census microdata relating to the 1990 and 1991 censuses in different countries. One spin-off of presenting this information to an international audience at this IASSIST meeting might be to stimulate discussion in this direction.

References

Campbell, D. (1987) 'The databank dossier', *New Statesman*, April 24.

Cullen, M. (1975) *The Early Victorian Statistical Societies*, Brighton: Wheatsheaf.

Denham, C.J. (1986) 'Census microdata in Great Britain: the possibilities', *Nutzung von anonymisierten Einzelangaben aus Daten der amtlichen Statistik: Bedingungen und Möglichkeiten*, Verlag W. Kohlhammer.

Gallup (1991) Gallup Political & Economic Index, Report No. 368, April.

Gates, G.W. (1988) Census bureau microdata: providing useful research data while protecting the anonymity of respondents, *Proceedings of the Social Statistics Section of the American Statistical Association Annual Meeting*, New Orleans, Louisiana, August 22-25.

Goyder, J., and Leiper, J. McK. (1985) 'The decline in survey response: a social values interpretation', *Sociology*, vol. 19, 1, pp. 55-71.

Griffin, R.A., Navarro, A., and Florez-Baez, L. (1989) 'Disclosure avoidance for the 1900 census', U.S. Bureau of the Census, paper prepared for presentation at the 1989 Joint Statistical Meetings, Washington, D.C., August.

Her Majesty's Government (1981a) 1981 Census of Population: Confidentiality and Computing, presented to parliament by the Secretary of State for Health and the Secretary of State for Scotland, Cmmd 8201, London: HMSO.

Her Majesty's Government (1981b) 'Initial study of the Office of Population Censuses and Surveys', Annex to the Review of the government statistical services, Cmmd 8236, London: HMSO.

Her Majesty's Government (1991) 1991 Census of Population: Confidentiality and Computing, presented to parliament by the Secretary of State for Health and the Secretary of State for Scotland, February, Cmmd 1447, London: HMSO.

Marsh, C., Arber, S., Wrigley, N., Rhind, D., and Bulmer, M. (1988) "The view of academic social scientists on the 1991 UK Census of Population: a report of the Economic and Social Research Council Working Group", *Environment and Planning A*, vol. 20:851-889.

Marsh, C., Skinner, C., Arber, S., Penhale, B., Openshaw, S., Hobcraft, J., Lievelesy, D., and Walford, N. (1991a) 'The case for samples of anonymised records from the 1991 census', *Journal of the Royal Statistical Society (A)*, vol. 154 (2), 1991.

Marsh, C., Dale, A., and Skinner, C. (1991b) "Safe data or safe settings: disseminating information from the British Census of Population", paper to be presented to

the ISI meetings in Cairo, September.

Norris, P. (1983) 'Microdata from the British Census', in David Rhind ed. *A Census User's Handbook*, London and New York: Methuen.

Redfern, P. (1976) Releasing statistics as aggregates (tables) or tapes of anonymised individual data (id). OPCS, London: mimeo, available from the author at 17 Fulwith Close, Harrogate, HG2 8HP.

Senior, M., and Cole, K. (1991) '1981 Census costs Department of Health £190,000!', *Manchester Computing Centre Newsletter*, no. 183, p.2.

¹ Presented at the IASSIST 91 Conference held in Edmonton, Alberta, Canada. May 14 - 17, 1991.

² The Federal Republic of Germany has a policy against the general release of microdata collected by the state. It will not allow its Labour Force Survey to be released for secondary analysis, for example. Since the Statistical Office of the European Commission takes the most restrictive policies of its member states as guidelines on release of microdata, this means that Europe-wide release of the Labour Force Survey is precluded.

³ In fact, the value of census data to academic researchers does not seem to decline as the census data becomes less timely; if usage of small area statistics is a guide, the usage of these increased throughout the 1980s.

⁴ Perhaps this is the reason why two other data sources collected compulsorily by law (the New Earnings Survey and Census of Employment) are only released in aggregates, albeit of very small units.

⁵ (Trans.) For others to thresh, or, more colloquially, let someone else work out what it all means! It is still the motto of the Royal Statistical Society.

⁶ Households with all members absent on census night and which fail to return a voluntary return to the census offices once they get back home.

⁷ The body which funds purchases and support of main-frame computing in British universities. It is soon to become a subcommittee of the main Universities Funding Council.