

**ESRC/JISC QUESTIONNAIRE:
VIEWS ABOUT THE 2001 CENSUS OF POPULATION**
Your chance to influence the 2001 Census

August 1997

Dear Census User,

Oh no, you say, another of those long questionnaires from a fellow academic! It looks like it will take at least 30 minutes to fill in. So, what's in it for me? Well, if you are a user of the Census of Population data or have thought the data might be useful in your research, this questionnaire gives you an opportunity to influence

- the content of the 2001 Census
- the concepts to be used in the 2001 Census
- the outputs from the 2001 Census
- the shape of the ESRC/JISC programme for the 2001 Census

The questionnaire is packed with up to date information about current proposals for the 2001 Census, which you may find useful in planning research activities in the next decade. *Section 1* covers the topics and questions used in the June 1997 Census test, which are a short list for the 2001 Census. We need to know which are vital to your research and whether the questions and codings are suitable. *Section 2* outlines the key changes in concepts proposed for the 2001 Census. We need to know your reactions to the proposals. *Section 3* asks briefly about your use of 1991 Census data, outlines some of the proposals for outputs from the 2001 Census, and asks for your opinions and suggestions. *Section 4* asks you about the kinds of services you want from the ESRC/JISC Census Programme.

So, take the questionnaire away with you on holiday or to a Conference, complete and send it back in the pre-paid envelope. Your views will be kept entirely confidential but they will be *essential* inputs to two reports:

- a report to ESRC/JISC recommending what data should be purchased from the 2001 Census and how they should be disseminated
- a report to the Census Offices making the "business" case for census topics and questions, and making recommendations about the outputs from the 2001 Census.

Thanking you in anticipation.

Philip Rees

Co-ordinator, ESRC/JISC Census Programme, 1992-97
Convenor, ESRC/JISC Workshops Planning for the 2001 Census, 1996-98

*QUESTIONNAIRE TO USERS OF CENSUS DATA:
VIEWS ABOUT THE 2001 CENSUS OF POPULATION*

Summer 1997

Philip Rees, School of Geography, University of Leeds

ESRC and JISC want to know your views

The ESRC and JISC are preparing the case for funding purchase and support of datasets from the 2001 Census of Population. ESRC and JISC wish to know how much use academics made of the 1991 Census of Population datasets, and the associated programme services. ESRC and JISC want to know what are your requirements for data from the 2001 Census and what sort of programme of support is needed. If you are unfamiliar with the ESRC/JISC Census Programme, it would be helpful to read a paper which provides an overview and is available on the Census Web Site at MIDAS. The URL is <http://www.midas.ac.uk/census/cenprog.doc>.

The Census Offices need to know about your requirements

The UK Census Offices¹ are currently preparing the business case for census questions. The final choice of questions will be announced in a White Paper in 1998, and are seeking views of their customer sectors, including the academic community. The Census Offices have consulted intensively over the past 18 months on the questions to be asked in the 2001 Census. In June 1997 a Test Census was carried out in selected areas of the country. A copy of the June 1997 Test Census form has been placed on the Census Web Site at MIDAS. The inclusion or format of a question in the 1997 Test does not necessarily signify acceptance for the 2001 Census and the Test will result in modifications to questions. It is still vital that strong cases are made for the questions currently proposed.

This questionnaire is organised in 4 sections:

1. Topics and questions proposed
2. Proposed changes in concepts for the 2001 Census
3. Proposed changes in outputs for the 2001 Census
4. The ESRC/JISC Census Programme for the 2001 Census

Please respond to each of the questions on which you have a view. If you wish to make a new proposal about a census question or output, please provide clear and cogent arguments about how the information would be used and why other sources cannot supply the data. Space is provided for your arguments but feel free to provide additional comments on separate sheets. The survey of census users carried out in 1987 was extremely influential². We aim to emulate its success.

Please return the completed questionnaire, using the pre-paid envelope, by **30 September 1997** to

Mrs. Christine Macdonald
School of Geography, University of Leeds, Leeds LS2 9JT
tel: 0113 233 6635, fax: 0113 233 3308, email: C.Mcdonald@geog.leeds.ac.uk

¹ There are three Census Offices: the Office for National Statistics, the General Register Office Scotland and the Northern Ireland Statistics and Research Agency.

² Cathie Marsh, Sara Arber, Neil Wrigley, David Rhind and Martin Bulmer (1988) Research Policy and Review 23. 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*, 20, 851-889.

1. TOPICS AND QUESTIONS PROPOSED FOR THE 2001 CENSUS

The Census Test carried out in June 1997 used a questionnaire with 11 household questions and 32 individual questions. The questions have been thoroughly debated over 18 months (1995-97) in the Census Offices' Content Working Group, which includes representatives of census customer sectors. Some questions are certain to be asked in the 2001 Census such as names of household residents, names of visitors and their usual address, sex, date of birth, and marital status; other questions will require a strong justification to be included. The text of the questions is Crown Copyright and has been reproduced here with permission from the Office for National Statistics. There are one or two small differences in the Census Test questions used in Scotland and Northern Ireland, which are noted below.

For each of the topics which have been used in the June 1997 Census Test schedule please **rate** how important each would be for your research and teaching.

Please **circle** one of the numbers 1-4 or 9 against each topic.

1. essential
2. highly desirable
3. of interest
4. very low priority
9. not a question of interest for my research/teaching

The topics that are certain to be asked (i.e. those listed above) have been omitted from this questionnaire and so some topic/question numbers will be missing. Space is provided for you to supply arguments about any particular topic. Please give your reasons, especially about any topics which you consider to be category 1. Please comment as well on the form of the question and the coding categories proposed.

HOUSEHOLD TOPICS

Please **circle** one of the numbers 1-4 or 9 against each topic.

Q	Topic	Essential	Highly desirable	Of interest	Very low priority	Not of interest
H1	Type of accommodation	1	2	3	4	9
H2	Sharing of accommodation	1	2	3	4	9
H3	Exclusive use of bath/shower/toilet	1	2	3	4	9
H4	Rooms	1	2	3	4	9
H5	Lowest floor level of accommodation (extended)	1	2	3	4	9
H6	Central heating	1	2	3	4	9
H7	<i>Garden or yard (new)</i>	1	2	3	4	9
H8	Cars or vans available	1	2	3	4	9
H9	Tenure	1	2	3	4	9
H10	Landlord	1	2	3	4	9
H11	Furnished/unfurnished accommodation	1	2	3	4	9

INDIVIDUAL TOPICSPlease **circle** one of the numbers 1-4 or 9 against each topic.

Q	Topic	Essential	Highly desirable	Of interest	Very low priority	Not of interest
I5	Student/schoolchild status	1	2	3	4	9
I6	Term-time address	1	2	3	4	9
I7	Country of birth	1	2	3	4	9
I8	Ethnic group	1	2	3	4	9
I9	<i>Religious group (new)</i>	1	2	3	4	9
L1	Language (Irish, Gaelic, Welsh)	1	2	3	4	9
I10	<i>Provision of unpaid help (new)</i>	1	2	3	4	9
I11	<i>General health (new)</i>	1	2	3	4	9
I12	Long-term limiting illness	1	2	3	4	9
I13	<i>Receipt of unpaid personal help (new)</i>	1	2	3	4	9
I14	Educational qualifications (extended)	1	2	3	4	9
I15	Usual address one year ago	1	2	3	4	9
I17	Employment status	1	2	3	4	9
I18	<i>Number of paid jobs (new)</i>	1	2	3	4	9
I19	<i>Years since paid job (new)</i>	1	2	3	4	9
I21	Main job	1	2	3	4	9
I22	Main things done in job	1	2	3	4	9
I23	Supervision/management responsibilities	1	2	3	4	9
I24	Hours worked in main job	1	2	3	4	9
I25	Organisation or company of main job	1	2	3	4	9
I26	What organisation or company makes/does	1	2	3	4	9
I27	<i>Number of people employed (new)</i>	1	2	3	4	9
I28	Address of workplace	1	2	3	4	9
I29	Mode of travel to work	1	2	3	4	9
I30	Other activities in last week	1	2	3	4	9
I31	<i>Total gross income (new)</i>	1	2	3	4	9

QUESTION WORDING AND CODING: HOUSEHOLD TOPICSH1 What type of accommodation does your household occupy?

Categories: A **whole** house or bungalow that is: Detached, Semi-detached, Terraced (including end-of-terrace); A flat, maisonette, or apartment that is: In a purpose-built block of flats or tenement; Part of a converted or shared house; In a commercial building; Mobile or temporary structure: A caravan or other mobile or temporary structure

H2 Is your household's accommodation self-contained, that is with all rooms, including the kitchen, bathroom and toilet behind a door that only your household can use?

Categories: Yes, only our household has access to our accommodation; No, we share access with another household

H3 Do you have a bathroom/shower and toilet for use only by your household?

Categories: Yes; No

H4 How many rooms do you have for use by your household?

Please write in the number of rooms.

H5 What is the lowest floor level of your household's living accommodation?

Categories: Basement or semi-basement; Ground floor (street level); First floor (floor above street level); Second floor; Third or fourth floor; Fifth floor or higher

H6 Does your accommodation have central heating?

Categories: Yes, in some rooms or all rooms; No

H7 Does your household's accommodation have a garden or yard? (new)

Categories: Yes, *not shared* with any other household; Yes, *communal*, or *shared* with at least one other household; No

H8 How many cars or vans are owned, or available for use, by one or more members of your household?

Categories: None; One; Two; Three; Four or more, *please write in the number*

H9 Does your household own or rent the accommodation?

Categories: Owns outright; Buying with a mortgage or loan; Pays part rent and part mortgage (shared ownership); Rents; In some other way

H10 Who is the landlord?

Categories: Council (Local Authority), Scottish Homes, Northern Ireland Housing Executive; Housing Association, Housing Co-operative, Charitable Trust; Private landlord or letting agency; Employer of a household member; Relative or friend of a household member; Other, *please write in.*

H11 Is the accommodation provided furnished or unfurnished?

Categories: Furnished or partly furnished; Unfurnished

Q: Do you have any comments on the wording or coding of the household questions?

Comments on questions or coding:

QUESTION WORDING AND CODING: INDIVIDUAL TOPICSRelationship within the household

Categories: Their relationship to person number 1. Husband or wife; Living as a couple; Son or daughter (including adopted); Stepchild; Brother or sister; Half/step brother or sister; Parent; Parent-in-law; Son-in-law or daughter-in-law; Brother-in-law or sister-in-law; Niece or nephew; Grandparent; Grandchild; Not related; Other related, *please write in*

I5 Are you a student or schoolchild in full-time education?

Categories: Yes; No

I6 Is this your term-time address?

Categories: Yes; No

I7 What is your country of birth?

Categories: England; Scotland; Wales; Northern Ireland; Irish Republic; Elsewhere, *please write in*

I8 To which one of these ethnic groups do you consider you belong?

Categories: White; Black-Caribbean; Black-African; Black-other, *please write in*; Indian; Pakistani; Bangladeshi; Chinese; Mixed ethnic group, *please write in*; Any other ethnic group, *please write in*

(The Northern Ireland question includes Irish Traveller and excludes Black-other)

I9 Do you consider you belong to a religious group? (new)

Categories: No; Christian; Buddhist; Hindu; Islam/Muslim; Jewish; Sikh; Any other religion, *please write in*

In Northern Ireland the question on religion is as follows:

I9 What Religion, Religious Denomination or Body do you belong to?

Categories: Roman Catholic; Presbyterian; Church of Ireland; Methodist; Other, *please write in*; None

L1 In Northern Ireland, Wales and Scotland there are likely to be Census questions on Irish/Gaelic/Welsh respectively, which will be closely modelled on the equivalent questions in the 1991 Census.

I10 Do you provide substantial unpaid personal help for a friend or relative with any long-term illness, health problem or disability? (new)

Categories: Yes, to someone in my household; Yes, to someone outside my household; No

I11 Over the last twelve months would you say your health has on the whole been: (new)

Categories: Good; Fairly good; Not good

I12 Do you have any long-term illness, health problem or disability which limits your daily activities or the work you can do?

Categories: Yes, including problems which are due to old age; No

I13 Do you receive substantial unpaid personal help from a friend or relative? (new)

Categories: Yes, from someone in my household; Yes, from someone outside my household; No

Q: Do you have any comments on the wording or coding of these individual questions?

Comments on questions or coding:

QUESTION WORDING AND CODING: INDIVIDUAL TOPICS (CONTINUED)I14 Have you obtained any of the qualifications listed?*Categories:*

- (1) No qualifications;
- (2) Level 1 NVQ/SVQ, Foundation GNVQ/GSVQ;
- (3) GCSE (grades D to G), CSE (grades 2 to 5), SCE O (grades D and E), SCE Standard (grades 4 to 7), SCOTVEC National Certificate Modules;
- (4) GCSE (grades A to C), GCE 'O' level passes, CSE grade 1, SCE O (grades A to C), SCE Standard (grades 1 to 3), School Certificate, Matriculation;
- (5) Level 2 NVQ/SVQ Intermediate GNVQ/GSVQ;
- (6) BEC (General), BTEC (General), City and Guilds Craft or Ordinary level, RSA Diploma;
- (7) GCE 'A' level, SCE Higher (grades A to C);
- (8) Level 3 NVQ/SVQ, Advanced GNVQ/GSVQ;
- (9) ONC/OND, TEC (National), BEC (National), BTEC (National), City and Guilds Advanced Craft or Final Level;
- (10) Level 4 NVQ/SVQ;
- (11) HNC/HND, BEC (Higher), TEC (Higher), BTEC (Higher), RSA Advanced Diploma;
- (12) First degree e.g. BSc, BA, BEd other degree-level qualification including MAs at first degree level;
- (13) Level 5 NVQ/SVQ;
- (14) Higher degree e.g. MSc, MA, PGCE, PhD;
- (15) Qualifications other than those listed *please give the title and grade, or level attained including any professional qualifications, for instance nursing, accountancy, teaching. If you have gained qualifications from outside the UK, give details including the name of the country where they were obtained.*

I15 What was your usual address one year ago?*Categories:* The address shown on the front of this form; Child aged under one; No usual address one year ago; Elsewhere, *please write in.**If you were a student or schoolchild one year ago, give the address at which you were living during term-time. For a child born after 15 June 1996, tick 'Child aged under one'.*I17 Last week, did you do any paid work as an employee or as self-employed/freelance?*Tick all the boxes that apply.**Categories:* Yes, working as a paid employee **full-time** (more than 30 hours a week); Yes, working as a paid employee **part-time** (30 hours or less a week); Yes, working as self-employed/freelance; No, not workingI18 Last week, how many paid jobs did you have as an employee or as self-employed/freelance? (new)*Categories:* One; Two; Three or moreI19 How many years is it since you had a paid job as an employee or as self-employed/freelance? (new)*Categories:* 10 years or less, *please write in number of years*; More than 10 years, or have never had a paid job**Q: Do you have any comments on the wording or coding of these individual questions?***Comments on questions or coding:*

QUESTION WORDING AND CODING: INDIVIDUAL TOPICS (CONTINUED)**I21 What is your main job?**

Give the full title, for example, 'office cleaner', 'secondary school teacher', 'Corporal-radio operator'. If you do not have a job title, for example if you work freelance, give a general description of your job.
Please write in below

I22 What are the main things you do in your job?

Please write in.

I23 Do you supervise any other employees or do you have any management responsibilities?

Categories: No; Yes, I am a supervisor or foreman; Yes, I am a manager

I24 How many hours a week do you usually work in your main job?

Please write in number of hours worked a week.

I25 What is the full name of the organisation or company you work for in your main job?

Categories: Work for an organisation or have my own business, *please write in the organisation's name*; Self-employed/freelance (I do not have my own business or company); Work for a private individual

I26 At the place where you work in your main job, what does the organisation you work for make or do? If you are self-employed/freelance, what do you make or do?

Please write in.

I27 How many people are employed by the organisation or company you work for, at the place you work for in your main job?

Categories: 1-9; 10-24; 25-499; 500 or more; Self-employed/freelance not employing others

I28. What is the address of the place where you work?

Categories: No fixed place; Mainly work at or from home; The address below

I29 How do you travel to work?

Categories: Work mainly at home; Underground, tube, metro or light rail; Train; Bus, minibus or coach (public or private); Taxi or minicab; Motor cycle, scooter or moped; Driving a car or van; Passenger in a car or van; Pedal cycle; On foot; Other *please write in.*

I30 Last week, in addition to any work, were you doing any of the following activities?

Tick either 'Yes' or 'No' for each activity.

Categories: Looking after the home or family; On a government training or employment scheme; Waiting to start a job that you had already accepted; Unemployed and looking for work; A student or in other full-time education; Unable to work due to long-term sickness or disability; Retired from paid work

Q: Do you have any comments on the wording or coding of these individual questions?

Comments on questions or coding:

QUESTION WORDING AND CODING: INDIVIDUAL TOPICS (CONTINUED)**I31** *What is your total gross income from all sources? (new)*

Tick the box for the range into which your income falls and count *all* income including Earned income, Pensions, Child benefit, Other state benefits, Interest, Rent, Other regular allowances, Any other sources

Categories:

Per week	Per year (approximately)
(1) Nil	Nil
(2) Less than £60	Less than £3,000
(3) £60 to £119	£3,000 to £5,999
(4) £120 to £199	£6,000 to £9,999
(5) £200 to £299	£10,000 to 14,999
(6) £300 to £479	£15,000 to 24,999
(7) £480 or more	£25,000 or more

Q: Do you have any comments on the wording or coding of this individual question? (I31)

Comments on questions or coding:

There is a small chance that one or two strongly supported questions could still be added to the Census questionnaire (e.g. Ability in English and Travel to Place of Study).

Q: Are there any other questions which you feel must be included in the 2001 Census?

Please provide an argued case for any question:

2. PROPOSED CHANGES IN CONCEPTS FOR THE 2001 CENSUS

A number of changes in the way in which the Census is collected and processed are currently proposed by the Census Offices. These would have a major impact on research uses of Census outputs after 2001. A number of new outputs have been suggested by the Census Offices and user groups. *It should be stressed that the proposed changes described here should not be regarded as official Census Office policy but rather as indications of policy options, the feasibility of which are being vigorously pursued.*

Please indicate your views about these changes by circling the appropriate number using the scale:

1. essential
2. highly desirable
3. of interest
4. very low priority
9. not of interest

Please add comments to support your requirements wherever possible.

2.1 THE POPULATION BASE FOR THE 2001 CENSUS

If you do not you have a research or teaching interest in the Population Base for the 2001 Census, please go to section 2.5 on the Samples of Anonymised Records.

The Census Offices have suggested that there should be a new population base for the preparation of outputs from the 2001 Census. This population base will count students at their term address as their usual residence. There are several reasons for doing this: (1) this will align the Census population base with that used for the mid-year district/unitary authority population estimates, which include students; (2) students are likely to be more numerous in 2001 than in 1991 and probably spend more time at their places of education than at their parental home (if different); and (3) difficulties in measuring student migration need to be addressed.

Q: How important is this change in the Census population base for your work?

Essential	Desirable	Undesirable	Very undesirable	Not of interest
1	2	3	4	9

One of the consequences of this change in population base would be a lack of direct comparability between 1991 and 2001 population bases. Are you likely to need some total population counts for areas for a variety of population bases? This information might take the form of extended versions of Table 1 in the Small Area Statistics, Local Base Statistics, County Monitors and County Reports, and a 1991 population base for children under 18 years of age.

Q: How important is provision of information from the 2001 Census on a 1991 population base?

Essential	Highly desirable	Of interest	Very low priority	Not of interest
1	2	3	4	9

Q: Do you have any comments on the Population Base proposal?

Comments:

³ A comprehensive account of current Census Office proposals is given in *Census News No.38, July 1997* which is available from ONS (tel 01329 813800) or GROS (tel 0131 314 4254).

2.2 THE ONE NUMBER CENSUS IN 2001

If you do not you have a research or teaching interest in Underenumeration in the 2001 Census, please go to section 2.3 on the Processing and Publication Timetable for the 2001 Census.

You will recall considerable and interest at the time of the 1991 Census in underenumeration, which has been estimated at about 2.2% (one million missing persons). A decision was taken not to re-base the 1991 mid-year estimates on the 1991 Census. This underenumeration was associated with difficulties in contacting hard-to-enumerate population, which occur with any census, and with the reluctance of residents concerned about the poll tax to be enumerated. The Census Offices have begun research on methods to impute missing individuals and missing households in 2001, using cross-checks with independent data from coverage surveys, capture-recapture techniques, administrative records, demographic analysis and a statistical model linking underenumeration to area characteristics, so that a single estimate of the population is produced for 2001. Research into these techniques is underway in the Census Offices and a project is being undertaken under Professor Ian Diamond’s direction at Southampton University⁴. If this research is successful before 2001, the Census Offices propose to publish only one set of population figures for 2001, the *one number census*, based on adding sufficient individual and household replicates, area by area.

Q: How important is the one number census proposal for your research?

Essential	Highly desirable	Of interest	Very low priority	Not of interest
1	2	3	4	9

Comments. Please add any advice on methods for producing a one number census.

⁴ Graham Jones (1997) Planning the 2001 census: only four years to go. *Population Trends*, 88, 31-35.

2.3 THE PROCESSING AND PUBLICATION TIMETABLE FOR THE 2001 CENSUS

If you do not you have a research or teaching interest in the Processing and Publication for the 2001 Census, please go to section 2.4 on Area Statistics from the 2001 Census.

The Census Offices are considering radical changes to the way in which forms are processed and the results published in the 2001 Census. The Census Offices are investigating the postback method for returning Census Forms, which promises considerable savings. If this method is adopted then the data on the forms will be input to computer databases on a first come, first served basis. No output processing (bar test runs of output systems) will occur until the Census returns are declared "complete". Further imputed returns could be added to produce a one number census (see above).

The basic proposal is to process data for the whole country before releasing any results⁵. There would be no earlier serial output and no preliminary counts of the type produced from previous censuses. This strategy would (1) allow the provision of national results as a priority and allow the development of an enhanced set of statistics at local level; (2) allow improved quality checks to be made and the results to be released with the statistical output; (3) allow for the introduction of a one number approach if this proves feasible; and (4) allow for free-flow processing (with the inclusion of late returns from the field). Both free-flow processing and a one number census approach would result in a national output base being available more or less at a single point in time.

In the consultation meetings about the 2001 Census the majority of users from the non-academic sector have stated that they prefer "one hundred percent correct" outputs from the Census on a longer timetable. Some academics have put forward the view that a planned sequence of version releases (cf. software releases) which are revised as a result of user experience will provide information sooner and that the "one hundred percent correct" goal is unrealistic. It is, however, likely that most of the structures of standard outputs (e.g. table frameworks, data formats) will be released well before the main outputs so that users can test their analysis software.

A number of different views have been put forward in consultation meetings about these proposals. Please indicate what your opinion is by circling the appropriate response in reaction to the statements set out below.

S: The processing and publication timetable suggested is a sensible plan for producing outputs from the 2001 Census.

Strongly agree	Agree	Disagree	Strongly disagree	Not of interest
1	2	3	4	9

S: I would like to see some preliminary outputs from the 2001 Census produced within one year of the 2001 Census.

Strongly agree	Agree	Disagree	Strongly disagree	Not of interest
1	2	3	4	9

S: The Census Offices should plan a sequence of releases of output data sets rather than delay production until the data are "one hundred percent correct".

Strongly agree	Agree	Disagree	Strongly disagree	Not of interest
1	2	3	4	9

⁵ This paragraph is reproduced with ONS permission from *Census News No.38*, p9.

3. PROPOSED CHANGES IN OUTPUTS FOR THE 2001 CENSUS

3.1 AREA STATISTICS FROM THE 2001 CENSUS

If you do not you have a research or teaching interest in Area Statistics from the 2001 Census, please go to section 3.2 on the Samples of Anonymised Records.

The term *Area Statistics* refers to tables of counts produced for geographical areas at a variety of scales. In 1991 there were two main sets: the Small Area Statistics (simpler tables), and the Local Base Statistics (detailed tables).

3.1.1 Use of Area Statistics from the 1991 Census

Please indicate which Census data sets you have used by circling appropriate numbers.

Area Statistics	
<i>Dataset</i>	
1991 Census Small Area Statistics (SAS)	1
1991 Census Local Base Statistics (LBS)	2
1991 Census Statistics Derived from SAS or LBS	3
Other Area Statistics (please specify) _____	4
<i>Scale</i>	
1991 Census SAS for EDs, OAs	5
1991 Census LBS for Wards/Postal Sectors	6
1991 Census SAS or LBS, for districts, counties	7
1991 Census SAS for Constituencies	8
<i>Country</i>	
1991 Census SAS or LBS, England and Wales	9
1991 Census SAS or LBS, Scotland	10
1991 Census SAS, Northern Ireland	11

Notes: EDs = enumeration districts; OAs = Output Areas

3.1.2 Proposals for output areas from the 2001 Census

Over the period end-1995 to start-1997, the nature of the geography to be used for Small Area Statistics has been thoroughly debated in the forum of the Census Offices' Output Working Group⁶ and investigated by different researchers^{7,8}. The following views have gained considerable support:

- a uniform geography across whole UK
- only one standard SAS set at the smallest scale
- output areas separated from collection areas
- output areas based on aggregations of unit postcodes (or addresses)
- output areas to fit wards current in 2001 (exact or best fit)
- look up tables linking output areas to other areas as these change after 2001
- look up tables linking output areas to 1991 Census areas
- output of digital boundaries concurrently with statistics
- the detail of output statistics adjusted to geographical scale.

⁶ Phil Rees (ed.) (1997) The debate about the geography of the 2001 Census: collected papers from 1995-6. *Working Paper 97/1*, School of Geography, University of Leeds.

⁷ David Martin (1997) From enumeration districts to output areas: experiments in the automated creation of a census output geography. *Population Trends*, 88, 36-42.

⁸ Phil Rees and Oliver Duke-Williams (1997) Experiments with and recommendations for the creation and release of Small Area Statistics from national censuses. Report prepared for the Statistical Disclosure Control Project, ESPRIT Programme of the European Union. Available from the School of Geography, University of Leeds.

These views reflect needs for both administrative and postal geographies. In the 1987 survey⁹ there was overwhelming support for postcode based small area geography but only some of this need was met in the outputs from the 1991 Census. Postal sector SAS and the ED/PC directory were produced in England and Wales, while in Scotland Output Areas were collections of unit postcodes. Output areas in 2001 could be designed to be of uniform size making comparisons of statistics more valid. They could be defined from the same geographic base being used for ED planning and designed so that confidentiality thresholds were all met. Their boundaries and postcode constitutions could be published prior to the production of census statistics. Statistics for larger areas will be produced by aggregating Output Area statistics, on either an exact fit or a best fit basis.

The consultations have ruled out the possibility that more than one set of small area statistics would be produced, because of the perceived risk of disclosure of information about identifiable individuals. Whatever decisions are taken about the form of output areas from the 2001 Census, there are likely to be directories or look up tables linking to various current and historic geographies.

The options for output areas therefore are as follows.

- a. Use enumeration districts defined for collection purposes in 2001. These would nest into wards.
- b. Use output areas built up from unit postcodes (as described above) that nest into wards and postal sectors on an exact or a best fit basis.
- c. Use enumeration districts or output areas from the 1991 Census again. This would maintain historical comparability.

Q: Please rate these output area options in terms of importance for your research and please provide a ranking by circling the appropriate number.

Option	Essential	Highly desirable	Of interest	Very low priority	Not of interest	Rank
2001 EDs	1	2	3	4	9	1 2 3
OAs built from postcodes	1	2	3	4	9	1 2 3
1991 EDs/OAs	1	2	3	4	9	1 2 3

Comments on output areas from the 2001 Census:

⁹ Marsh *et al.* (1988), op. cit.

3.1.3 Addresses and associated geographical codes in the 2001 Census

One major development in geographic referencing since the 1991 Census has been the creation of the Address Point product by Ordnance Survey Great Britain and an equivalent product by Ordnance Survey Northern Ireland. Each address is associated with a one metre accurate grid reference. The General Register Office Scotland continues to maintain its property and postcode digital Geographical Information System. Grid references are provided for the geographic centres of unit postcodes. The Census Offices are preparing GIS systems incorporating this information and 1991 Census digital boundaries as a means of planning the census collection operation in 2001. Enumerators will be provided with address lists and digitally produced maps for enumeration areas. These GIS databases are potentially available for use in outputs as well. Address Point, for example, can be used for the automatic production of unit postcode boundaries and contiguities needed in output area design¹⁰. Output areas could be defined automatically using address based building blocks rather than unit postcodes. However, use of Address Point and equivalent products may involve additional royalties for users and it is vital that a strong case is made for their use.

Please describe what uses you are likely to make of output areas which use either postcodes or addresses as the units from which they are built, if this option is adopted.

Uses of postcode or address based output areas:

3.1.4 Temporal comparability of area statistics between 1991 and 2001

One of the most important functions of a census is to measure population change. This has always been quite difficult using British censuses because the definitions of administrative, electoral, postal and census collection areas change for very good reasons. Many solutions have been proposed: grid cells¹¹, aggregating areas from two censuses to zones with common boundaries (produced from the 1981 Census), freezing an output geography for one census and using it in subsequent censuses¹², use of look up tables that link the geography at one point in time to another (as currently implemented via twice yearly updates of the ED/PC directory by ONS) or designing current output areas to nest into previous units, as 1991 OAs in Scotland nested within the 1981 ED.

Q: How useful would the following solutions to the time comparability problem be for your research?

Proposal	Essential	Highly desirable	Of interest	Very low priority	Not of interest
Look up tables linking 2001 OAs and grid cells	1	2	3	4	9
Look up tables linking 2001 OAs and 1991 EDs/OAs	1	2	3	4	9
Census Tracts for 2001 OAs and 1991 EDs/OAs	1	2	3	4	9
2001 OAs that aggregated to 1991 EDs/OAs	1	2	3	4	9
1991 EDs/OAs as 2001 output areas	1	2	3	4	9

Comments:

Please feel free to make other suggestions for solutions to the problem.

¹⁰ Martin (1997), op.cit.

¹¹ Bracken and Martin (1995), op.cit.

¹² Daniel Dorling (1995) *A new social atlas of Britain*. John Wiley, Chichester, Sussex.

3.1.6 Disclosure control

The contents of the 2001 SAS have not yet been discussed in detail, and their content will be determined through consultations in 1998 and 1999. They would take into account new topics, new codings and the 100% processing of census returns in 2001. More detailed tables resembling the 1991 Census Local Base Statistics could be published for larger areas (wards, local government areas). The statistics will be subject to anti-disclosure measures: these would include minimum thresholds for the number of persons and households (as in 1991), but the Census Offices are investigating secondary protection measures (in their “belt and braces” approach), and would be interested in user preferences, given the difficulties experienced by users in handling randomly perturbed cell counts.

Q: Please rate the acceptability of following protection measures for your research.

Proposal	Highly accepta ble	Accept able	Indiffer ent	Un- accept- able	Not of interest
Small random changes to cell counts (-1, 0, +1)	1	2	3	4	9
Rounding of cell numbers (e.g. to base 3 or base 5)	1	2	3	4	9
Swapping of a small number of records between areas	1	2	3	4	9
Suppression of cells with low counts	1	2	3	4	9

Comments:

Please feel free to make other suggestions for solution to the problem.

3.2 SAMPLES OF ANONYMISED RECORDS FROM THE 2001 CENSUS

If you do not have a research or teaching interest in the Samples of Anonymised Records from the 2001 Census, please go to section 3.3 on the Migration Statistics from the 2001 Census.

Please indicate which 1991 Census data sets you have used by circling appropriate numbers.

Microdata: Samples of Anonymised Records	Yes
1991 Census 1% Household File, Great Britain or Northern Ireland	1
1991 Census 2% Individual File, Great Britain or Northern Ireland or UK	1

By spring 1998 we need to have a clear view on the requirements for SARs from the 1991 Census. The Samples of Anonymised Records were a new product from the 1991 Census, which were commissioned by ESRC as a result of the work of Cathie Marsh and other members of the ESRC's working party on the 1991 Census. SARs are special outputs which have to be requested and paid for. Details of thinking and consultations to date are given in the *SARs Newsletter*, No.10 (August 1997)¹³. The proposals which are currently being discussed are as follows.

1.. The *1% Household SAR* is likely to be broadly similar to that in 1991, with addition of any new questions and minor modification of codings. This hierarchical file is a sensitive data set with respect to confidentiality and adding further detail is unlikely to be possible.

2. Several proposals for development of the *Individual SAR* have been made in meetings.

- Addition of attributes of the household an individual belongs to.
- A finer specification of the geography: Local Authority (Unitary Authority/District)
- A finer specification of the geography: sub-Local Authority (groups of wards)

These proposals might need expansion of the sample size from 2% to 4 or 5% and a reduction in detail in some variables (such as occupation and industry). The threshold population for SAR areas would need to drop from 120,000 to 40,000 to make possible publication for LA areas. As a result the suggestion has been made that a third *Geographical SAR* be defined.

Q: Please rate the following 2001 SAR proposals in terms of importance for your own research.

Proposal	Essential	Highly desirable	Of interest	Very low priority	Not of interest
1% Household (hierarchical) SAR	1	2	3	4	9
2% Individual SAR with more derived household variables	1	2	3	4	9
Individual SAR with LA geography but less variable detail	1	2	3	4	9
Individual SAR with sub-LA geography but with even less variable detail	1	2	3	4	9

Comments:

Please feel free to make other suggestions for SARs in 2001:

¹³ Available from the Census Microdata Unit, Faculty of Economic and Social Studies, University of Manchester, Manchester M13 9PL. Contact Ruth Durrell on 0161-275-4721.

3.3 MIGRATION STATISTICS FROM THE 2001 CENSUS

If you do not you have a research or teaching interest in the Migration Statistics from the 2001 Census, please go to section 3.4 on the Journey to Work Statistics from the 2001 Census.

Please indicate which 1991 Census set of Migration Statistics you have used by circling appropriate numbers.

Migration Statistics	Yes
1991 Census National Migration Tables, Great Britain (published volumes)	1
1991 Census Regional Migration Statistics, Great Britain (computer readable tables)	1
1991 Census Local Base Statistics, Great Britain (migration tables)	1
1991 Census Small Area Statistics, Great Britain (migration tables)	1
1991 Census, Special Migration Statistics, Great Britain, Wards/Postal Sectors (ONS Set 1)	1
1991 Census, Special Migration Statistics, Great Britain, Districts (ONS Set 2)	1
1991 Census Special Migration Statistics, Great Britain, Districts (Leeds re-estimates)	1
Other Area Statistics (please specify) _____	1

To date there has been little discussion of the nature of Migration Statistics to be produced from the 2001 Census. However, the difficulties encountered with the 1991 Census Migration Statistics need to be overcome. These include late delivery of the Special Migration Statistics (SMS), uncertain specifications and formats for the Regional Migration Statistics, lack of out-migrant information in the migration tables in the Area Statistics, and suppression of small flows in the SMS.

The One Number Census and the Processing proposals (see sections 2.2 and 2.3 above) provide an opportunity for a major improvement in the Migration tables in the Local Base Statistics and Small Area Statistics. In 2001 these can be designed to report out-migrants from areas as well as in-migrants so that researchers can look at intra-national migration balances by a wider set of characteristics (e.g. age-sex, ethnicity, activity and employment status, housing tenure, household attributes, occupations). The Special Migration Statistics have proved very valuable but suffer from the major disadvantage that small flows are suppressed for most tables.

Q: How important are the Migration Statistics for your own research?

Essential	Highly desirable	Of interest	Very low priority	Not of interest
1	2	3	4	9

Please provide comments and suggestions for the production of Migration Statistics from the 2001 Census.

Comments and suggestions:

3.4 WORKPLACE STATISTICS FROM THE 2001 CENSUS

If you do not you have a research or teaching interest in the Workplace Statistics from the 2001 Census, please go to section 3.5 on the Inclusion of the 2001 Census in the Longitudinal Study.

Please indicate which Workplace Statistics from the 1991 Census you have used by circling appropriate numbers.

Workplace Statistics	Yes
1991 Census Transport and Workplace National Tables (published volumes)	1
1991 Census Local Base Statistics (transport tables)	1
1991 Census, Special Workplace Statistics, Great Britain, Wards/Postal Sectors (Set A: residence zones)	1
1991 Census, Special Workplace Statistics, Great Britain, Wards/Postal Sectors (Set B: workplace zones)	1
1991 Census Special Workplace Statistics, Wards/Postal Sectors (Set C: flows)	1
Other Workplace Statistics (please specify) _____	1

To date there has been little discussion of the nature of Workplace Statistics to be produced from the 2001 Census. However, some modifications may be necessary if the Census Offices carry out their intention to process workplace information for all Census returns (100% processing). In the 1991 Census Workplace Statistics no additional protection measures were applied because only 10% of returns were processed and this sampling was regarded as sufficient protection for the data.

Q: How important are the Workplace Statistics for your own research?

Essential	Highly desirable	Of interest	Very low priority	Not of interest
1	2	3	4	9

Please provide comments and suggestions for the production of Workplace Statistics from the 2001 Census.

Comments and suggestions:

3.5 THE LONGITUDINAL STUDY: INCLUSION OF THE 2001 CENSUS

If you do not you have a research interest in the Longitudinal Study (England and Wales), please go to section 3.6 on Boundary Data from the 2001 Census.

Please indicate which Longitudinal Census data sets you have used by circling appropriate numbers.

Microdata: Longitudinal Study	
1971-1981 Linked Census Data	1
1981-1991 Linked Census Data	2
1971-1981-1991 Linked Census Data	3
Other LS Data (e.g. births, deaths, cancers), please specify	4

The Longitudinal Study links individuals with one of four birthdays from the 1971 Census through the 1981 Census to the 1991 Census of England and Wales. The Sample (roughly 1.05%) is refreshed by new births and immigrants and decays through deaths and immigration. The LS data remain on ONS computers and a table/analysis service is provided to academics via the LS Support Programme funded by ESRC/JISC at City University. A case must be made for the addition of a new set of records from the 2001 Census.

Q: How important is the Longitudinal Study for your own research?

Essential	Highly desirable	Of interest	Very low priority	Not of interest
1	2	3	4	9

Please provide comments and suggestions for the linkage of the 2001 Census to the Longitudinal Study.

Comments on the LS Microdata/Suggestions for 2001 Outputs:

3.6 BOUNDARY DATA FROM THE 2001 CENSUS

If you do not you have a research interest in the Boundary Data from the 2001 Census, please go to section 3.7 on Look Up Tables from the 2001 Census.

Please indicate which 1991 Census Boundary data sets you have used by circling appropriate numbers.

Boundary data	
1991 Census Digital Boundary Data, England and Wales	1
1991 Census Digital Boundary Data, Scotland	2
1981 Census Digital Boundary Data, Great Britain	3
Other Digital boundary data (please specify) _____	4
OS Digital Maps as background	5
OPCS/GROS ED planning maps (microfilm or paper copies)	6

Please indicate at which scales you have used boundary data.

Scales used:	
Enumeration districts/Output Areas	1
Wards/Postal Sectors	2
Districts	3
Counties/Scottish Regions	4
Regions (Standard or Government Office)	5
Other zones (please specify) _____	6

Boundary data in digital form is provided by the UKBORDERS service at the University of Edinburgh. The ED boundaries for England and Wales and the OA boundaries for Scotland are held in a large ARC/INFO and INGRES database from which user extractions can be generated at a variety of scales and formats. The UKBORDERS service is currently adding libraries of pre-selected boundaries commonly used. It is hoped that in 2001 that arrangements can be made to add Northern Ireland boundaries.

Q: How important are the Boundary Data for your own research?

Essential	Highly desirable	Of interest	Very low priority	Not of interest
1	2	3	4	9

Comments on the Digital Boundary Data/Suggestions for 2001 Outputs:

3.7 LOOK UP TABLES FOR THE 2001 CENSUS

If you do not you have a research interest in Look Up Tables for the 2001 Census, please go to section 3.8 on Flexible Outputs from the 2001 Census.

Please indicate which 1991 Census Look Up Tables you have used by circling appropriate numbers.

Look Up Tables	
1991 Census Area Master File	1
1991 Census ED/1991 PC Directory, England and Wales	2
1991 Census ED/post-1991 PC Directory, England and Wales	3
1991 Census Indexes, Scotland	4
1991 Census Ward/Functional Region, Great Britain	5
1991 Census Ward/Localities, Great Britain	6
1991 Census ED/1981 Census Ward, Great Britain	7
Other Look Up Table (please specify) _____	8

Since the 1991 Census was taken, the importance of having good Look Up Tables which relate one small area geography to another has become very clear. One of the most used products from the 1991 Census has been the Enumeration District/Postcode Directory which enables users to make estimates of SAS data for postal zones or to aggregate data georeferenced by unit postcode to the census geography. The importance of updating such Look Up Tables has also been realised and the ESRC/JISC Census Programme now purchases an annual update of the ED/PC Directory. Users have also worked hard to develop their own Look Up Tables and these have been deposited with the Census Dissemination Unit at Manchester Computing for general use. Research is currently under way on placing Look Up Tables in a more general relational database framework.

Q: How important are the Look Up Tables for your own research?

Essential	Highly desirable	Of interest	Very low priority	Not of interest
1	2	3	4	9

Comments on the Look Up Tables/Suggestions for 2001 Outputs:

3.8 FLEXIBLE OUTPUTS FROM THE 2001 CENSUS

If you do not you have a research interest in Flexible Outputs from the 2001 Census, please go to section 4 on ESRC/JISC Census Programme services.

Please indicate whether you have used Special Tabulations from the 1991 Census and how you funded them.

Special Tabulations used: Funding source:
--

In the consultations about the 2001 Census outputs that have taken place, users have expressed a strong preference for being able to develop their own tables from the Census or for being able to carry out their own statistical analysis using individual or household records. This demand has been met in two ways from the 1991 Census: through use of the Longitudinal Study and release of the SARs (see sections 3.2 and 3.5).

In the past such needs have been met through special tabulations by the Census Offices, which have been quite expensive. They have been used sparingly by academic researchers. Several technical developments are needed to improve this situation. The first is the use by the Census Offices of faster and easier to use tabulation software. A wide range of such packages are available. GROS have installed such a package and ONS are in discussion with a supplier. The second area of development need is to find solutions to the risks of disclosure presented by a greater flexibility in output.

The Census Offices are keen to meet user needs and arrangements similar to those agreed for the Longitudinal Study in England and Wales may be possible for the whole UK Census in 2001. The next question tries to gauge the demand for such a flexible output system based on the full Census database.

Q: How useful would a flexible census output system be for your research?

Essential	Highly desirable	Of interest	Very low priority	Not of interest
1	2	3	4	9

Comments: Please feel free to make suggestions.
--

4. ESRC/JISC CENSUS PROGRAMME SERVICES

The different data sets discussed in section 3 have been disseminated by the four Census Programme Units:

CDU	The Census Dissemination Unit, Manchester Computing, University of Manchester
CMU	The Census Microdata Unit, CCSR, University of Manchester
LSSP	The Longitudinal Study Support Programme, SSRU, City University
UKB	The UKBORDERS Service, Data Library, University of Edinburgh

The Programme has been co-ordinated by Phil Rees, School of Geography, University of Leeds.

We wish to receive feedback on the service that the Units and Co-ordinator offer academic sector census users. General questions are asked here. Some Units are conducting their own more detailed customer satisfaction surveys (e.g. the CMU are currently surveying users of the SAR datasets). In assessing the quality of service provided a four category scale is used:

1. Outstanding
2. Good
3. Satisfactory
4. Unsatisfactory
9. Service not used

Please **circle** the appropriate response.

The main purpose of this section of the questionnaire is to learn how services can be improved in future, balanced, of course, against the cost of providing them.

4.1 REGISTRATION PROCEDURES

4.1.1 Registration procedures for 1991 Census data sets

Although census data are provided free at the point of use, the agreements with the Census Offices require ESRC to obtain the signature of the individual user, who assents to the agreed terms and conditions. Many agreements require that institutions also register and take responsibility for user conduct. Simpler forms are available for registering classes of students. Registration documents are now provided on-line or via the Web. Registrations that apply to data held at more than one site are exchanged via the network..

Q: How satisfactory have you found registration procedures for 1991 Census data sets?

Outstanding	Good	Satisfactory	Unsatisfactory	Service not used
1	2	3	4	9

Comments on registration procedures:

4.2.1 Registration procedures for 2001 Census data sets

Proposals for safe data sets. The aim in negotiating agreements for 2001 Census data is to design *aone stop* registration procedure which subsumes all data sets released as safe (including previous census data). Institutions would sign one agreement. Individuals would sign one general agreement and then be able to use all census data sets. The on-line Census Unit services may be linked into the authentication system being developed by JISC for all national services. As much of the documentation and metadata about census data will

be moved to freely accessible and linked census Web sites where small taster data sets will be displayed (by agreement with the Census Offices and other suppliers).

Proposals for sensitive data. Another procedure will probably be needed for data sets which have to remain in the safe environment of the Census Offices and are accessed via Census Units (e.g. Longitudinal Study or a more general flexible outputs service). These data are more sensitive and require stronger agreements. They are also more expensive to service and the project using LS data need to be evaluation and approval.

Please remember that census data are *free at the point of use* for academic research or teaching purposes!

Q: How satisfactory do you think these registration procedures for 2001 Census data sets will be?

Outstanding	Good	Satisfactory	Unsatisfactory	Service not used
1	2	3	4	9

Comments on proposed registration procedures:

4.2 HELP WITH CENSUS DATA

Help is vital for researchers no matter what their level of expertise. Help to census users is provided through conversations and interactions with Census Unit personnel via phone, email and letter. It is also provided via documentation provided in registration packs, dataset manuals, software manuals, in computer files or via on-line information systems (using gopher, lynx, http or searchable database formats).

Q: How would you rate the help provided by the Census Programme?

Outstanding	Good	Satisfactory	Unsatisfactory	Service not used
1	2	3	4	9

Comments and suggestions on the help/advice services provided by the Census Units

4.3 DOCUMENTATION AND TRAINING

This comes in several forms:

1. *Census Office Publications.* In their Statistical Volumes, the Census Offices provide information on definitions and classifications. For the machine readable data sets, the Census Offices produce User Guides.
2. *Published Books.* The ESRC/JISC Census programme helped fund *The 1991 Census User's Guide* (edited by Angela Dale and Catherine Marsh, HMSO, London, 1993), the *Census Users' Handbook* (edited Stan Openshaw, GeoInformation International Cambridge, 1995) and the *Longitudinal Study 1971-1991* by Lin Hattersley and Rosemary Creeser (HMSO, London, 1995).
3. *Census Programme Publications.* Each Census Unit produces *Registration Packs* and *User Guides* which provide basic information about data sets. More detailed information is provided in *Codebooks and Glossaries* and *Manuals associated with software interfaces*. Several *training manuals, tutorial systems* and *example analyses* were produced by projects at Leicester University and Southampton University.
4. *Census Office and Census Unit Newsletters.* Short articles describe new data sets or new analyses.
5. *On-line Information.* Some of the published information contained in the above documentation is provided in on-line form (subject to copyright) and great deal more detail about data subsets, derived data, user supplied enhancements and the 2001 Census is provided on-line via Census Unit servers or Web sites.
6. *Searchable databases.* These can be use to interrogate to find data sets, variables or crosstabs of interest. (Examples are the LS Data Dictionary, the Data Archive's Biron System, and Williamson's MetaC91 for SAS and SARs variables).
7. *One day Workshops or Seminars covering particular data sets.* The Census Units organise a series of workshops each year either held at their centres or at user institutions upon request.
8. *Week long courses covering a wide range of data sets and topics.* Two week long courses were organised by the CMU and CDU for teachers and researchers that covered a range of different census data sets.

Q: How useful have the different forms of documentation/training been for your research using the 1991 Census?

Please *circle* the appropriate numbers.

Documentation forms	Essential 1	Highly desirable 2	Of interest 3	Very low priority 4	Not of interest 9
Census Office Publications	1	2	3	4	9
Published Books	1	2	3	4	9
Census Programme Documents and Packs	1	2	3	4	9
Newsletters	1	2	3	4	9
On-line Information	1	2	3	4	9
Searchable Databases	1	2	3	4	9
One Day Workshops	1	2	3	4	9
Week Long Seminars	1	2	3	4	9

Comments and suggestions on documentation and training for the 2001 Census:

4.4 METHODS OF ACCESS

The machine readable census data sets are stored on Census Office or Census Unit servers and are disseminated from there in a variety of ways:

Q: How useful have the following methods of access been for your research using the 1991 Census?

Please *circle* the appropriate numbers.

Methods of access	Essential	Highly desirable	Of interest	Very low priority	Not of interest
<i>Transfer of raw data to local computer</i>					
DAT tape	1	2	3	4	9
Other forms of magnetic tape	1	2	3	4	9
diskette	1	2	3	4	9
CD ROM/DVD	1	2	3	4	9
File transfer	1	2	3	4	9
<i>On-line to server</i>					
Login session	1	2	3	4	9
Web session	1	2	3	4	9
<i>Customised tables (e.g. LS)</i>	1	2	3	4	9

Comments and suggestions on methods of access for the 2001 Census:

4.5 SOFTWARE FOR EXTRACTION AND ANALYSIS OF CENSUS DATA

Data from the 1991 Census were output by the Census Offices as simple ASCII files in fixed record formats using spaces, tabs or commas as field delimiters. Paper documentation accompanied the data files which described data structures and provided data definitions. The job of users or Census Unit staff was to load these data into tabulation, analysis or display packages. These packages were of two kinds: bespoke software for handling census data and general purpose packages in wide use. Discussion of the nature of interfaces to Census data in the Second ESRC/JISC Workshop in April 1997 arrived at the following recommendations.

1. Census data should be output in spreadsheet and database formats, as well as in formatted text files.
2. Meta-data should be in computer readable form and made available with release of the data sets.
3. Extraction and analysis should be seamlessly linked as far as possible.
4. Batch processing of extractions, in which the user looks up filenames or variable codings, should be replaced by interactive software accessing full menus of this meta-information.
5. The World Wide Web should be used wherever possible as the standard interface to datasets, because this was an open system with standards accessible to virtually all academic users.
6. For the largest and most complex datasets, a twin track strategy should be pursued of improving the general interfaces and of producing simple extractions in high demand.

In order to plan for the provision of such interfaces and software ESRC/JISC need to know what kinds of software people wish to use to analyse and display census data.

Q: Please indicate which software packages are important for your research using census data?

Please provide information on software for data sets you have used or are likely to use.

Dataset	Software package and associated data set
<i>Area Statistics</i>	<div style="border: 1px solid black; height: 50px;"></div>
<i>SARs</i>	<div style="border: 1px solid black; height: 50px;"></div>
<i>Interaction Statistics</i>	<div style="border: 1px solid black; height: 50px;"></div>
<i>Longitudinal Study</i>	<div style="border: 1px solid black; height: 50px;"></div>
<i>Boundary Data</i>	<div style="border: 1px solid black; height: 50px;"></div>

4.6: SHAPE OF THE ESRC/JISC CENSUS PROGRAMME AFTER 2001

The ESRC/JISC 1991 Census of Population Programme included the following set of important sub-programmes.

1. Census Training Programme
2. Census Development Programme
3. Census Research Programme
4. Census Seminar Programme
5. Census Programme Co-ordination

The Fourth and Final ESRC/JISC Workshop will discuss ideas about these elements, but if you have ideas about their shape after 2001 please use the remaining space for comments.

BACKGROUND CHARACTERISTICS

Please indicate your employer or institution you are studying at.

- University 1 Other Educational Institution 2
- Other Research Institution 3
- Other (please specify) 4 _____

Please indicate your occupation and discipline by circling the relevant numbers.

<i>Occupation</i>		<i>Discipline</i>		<i>Discipline</i>	
Professor	1	Business	1	Mathematics	13
Reader	2	Community Health	2	Medical Science	14
Senior Lecturer	3	Computing	3	Social Policy	15
Lecturer	4	Demography	4	Social Research	16
Senior Research Fellow	5	Econometrics	5	Social Statistics	17
Research Fellow	6	Economics	6	Social Work	18
Research Assistant	7	Employment	7	Sociology	19
Programmer	8	Environment	8	Statistics	20
Teaching Assistant	9	Epidemiology	9	Survey Research	21
Technician	10	Ethnic Research	10	Transport	22
Postgraduate Student (PhD)	11	Geography	11	Urban Research	23
Postgraduate Student (Masters)	12	Health Policy	12	Other (please specify)	24
Undergraduate Student	13				
Other (please specify)	14				

Please use the remaining space to make any additional comments.

Thank you for completing the questionnaire.

ECONOMIC AND SOCIAL RESEARCH COUNCIL (ESRC)
JOINT INFORMATION SYSTEMS COMMITTEE (JISC)

*QUESTIONNAIRE TO USERS OF CENSUS DATA:
VIEWS ABOUT THE 2001 CENSUS OF POPULATION*

QUESTIONNAIRE COMMENTS

Philip Rees
School of Geography, University of Leeds

December 1997

This document reports the comments made by census users responding to the survey of Views about the 2001 Census of Population. This survey was carried out in the summer of 1997 by the School of Geography of the University of Leeds. Sets of five questionnaires were sent to each social science department in all UK Higher Education Institutions and to those independent research institutes recognised by ESRC. Questionnaires were also sent to individuals on lists of users maintained by the ESRC/JISC Census Programme Data Units and Co-ordinator. The questionnaire was distributed in the week beginning the 11th August 1997. Return of completed questionnaires was requested by the 30th September. This deadline was extended to the 30th November to cater for census users who had been missed by the first mailouts. James Harris of the Census Dissemination Unit made the questionnaire available for on-line completion on the Census Programme Web site: <http://census.ac.uk/>.

The target population at which the survey was aimed was the set of users of computer readable census data in all its forms. Since completion was voluntary and the lists of user departments and users incomplete, the returns cannot be regarded as a statistically representative sample of the target population. They probably represent the most active and knowledgeable census users in the academic community. However, these are the very views which need to be taken very seriously by both ESRC/JISC and the UK Census Offices.

The material is organised under the major topics upon which users were asked to comment. Comments under each topic are listed by respondent number. Each returned questionnaire was assigned a unique number on return. Questionnaires 132 to 140 were completed via the interactive Web version. All questionnaires were completed anonymously. The bulk of data entry of these comments was carried out by Li Heming (postgraduate student) and by Mrs Christine Macdonald (secretary) at the School of Geography, University of Leeds. The comments have been slightly edited where spelling required correction and where clarification was needed. A parallel data file was also prepared containing the responses to numeric coded questions in the survey.

The document is divided into four sections corresponding to the structure of the questionnaire.

1. Topics and questions proposed
2. Proposed changes in concepts for the 2001 Census
3. Proposed changes in outputs for the 2001 Census
4. The ESRC/JISC Census Programme for the 2001 Census

Each section is divided into subsections.

Part 1:

TOPICS AND QUESTIONS PROPOSED

1.1 Household questions

Questionnaire instructions:

Do you have any comments on the wording or coding of these household questions?

007: Define room size as before.

009: H8 If gross income is not asked in 2001 perhaps one way of estimating income or living standards would be to include the registration (year) of the vehicles. This would allow more sensitive analysis of deprivation, therefore, in rural areas where the ownership of any car maybe a necessity.

018: I would like to see an extra question on the provision of parking facilities following H8, say, H8a: Does your accommodation have parking facilities? Yes; No. H8b: What types of parking facilities are you using for your household vehicles? private drive; private-owned garage; hired garage; off-street parking; resident only off-street parking.

019: Question H6 - What is the definition of 'central' heating? for example, do storage heaters count as 'central heating'? They may be in some or all rooms but the heating does not come from 'central' source.

023: H4 - Definition of 'room'? H6 - Would like to know if central heating is usual. H7 - Does not distinguish between a 10 acre park and 6 by 10 feet patch of bare earth!

027: These seem comprehensive. More use of household characteristics of individuals needed in 2001 Census (e.g. the Individual SAR, or SMS).

028: Could H9 not deal with sub-letting or renting out living with a friend who owns/pays the rent, important for social exclusion.

034: H3: Even though bath/shower and toilet question have now been combined, this question is no longer relevant to the UK and should be dropped. It is no longer an indicator for poverty as it was in former days. H5: This question should be asked throughout UK not just in Scotland, particularly since children in flats was used by the Department of the Environment. H11: I don't see the need for this question.

039: H8: Apart from car-availability, ownership might be useful.

041: H6 I prefer coding: Yes/Some/None.

044: H8 Very convoluted delete 'one or more' 'Available for use' very legalistic term I consider (in normal usage) any taxi 'available for use'. If you mean company car, say so.

045: H2 - Access: why not merely avoid the word. H10 Housing Associations and Co-operatives may be Charitable Trusts.

047: H4 - The 'small kitchen' problem needs to be avoided in 2001 i.e. the 'number of rooms' should include even 'kitchenettes' if they're separate, to avoid misleading comparison between areas with different traditions.

048: H6 - If yes, do you use it as your main form of heating? (to investigate the suspected many tenants who cannot afford to use their built-in central heating.)

051: I10 This is a interesting area, as the numbers of people who are in this situation are unknown. However, I think it is likely to suffer from the usual problems of self-reported data. It will of course be interesting to

match answers to this against answers to I13! I11 Again, I'm a little sceptical about the quality of responses to the question - does hypochondria vary from place to place?

- 058: H4 - Bathrooms/toilets/laundry? Definition.
- 076: H4 - What is a room here? See H2? Shared toilet?
- 079: H4 - It would be useful to have the number of bedrooms separately. - Does total number of rooms include toilets/kitchens/bathrooms?
- 080: H10 - Delete 'or letting agency' term. Landlord may need classification - i.e. whoowns the property.
- 086: H5 - I am not sure anything useful will come out of this. What of houses built on steep slopes? Houses deliberately built 'upside down' etc. H1 - I hope this will be worded clearly on the Census form.
- 102: H4 - Are garages counted as rooms?
- 105: H6 - Need to know type of central heating (e.g. gas, electric...).
- 111: How will rooms be defined this time - e.g. are kitchens and bathrooms now included in H4? Should H4 read "...have for use only.." to reduce double counting, though this would be unnecessary for non-shared accommodation.
- 116: H3: bathroom/shower and toilet - separate questions would give more reliable responses. H8: available for use 'on most working days'. H9: Problem of ownership by a parent (students) or a relative, e.g. sibling through inheritance. *Suggestion* Can the census tackle dual+ residences? In the past year have you lived at another address for more than 52 days, e.g. at weekends or on working days.
- 119: Household questions should somehow take into account local variations. For example H5 is prone to misleading conclusions when comparing affluent and deprived areas. Same with H7 where communal and shared should form different categories.
- 121: In recently trying to determine (H4) proportion of households which are overcrowded, I came to the conclusion it would be more useful to know number of roomsexcluding kitchens (at any size), bathrooms and utility rooms.
- 128: Q18 must stipulate an Irish category not just as a "write-in".
- 130: H6 - This seems to be skewed as an indicator of housing quality by its provision in public sector housing which might otherwise appear poorer quality.
- 139: Is it going to be clear who is a household member and what counts as a room? Continuity of definitions with previous census would be appreciated.

1.2. Individual questions

Questionnaire instructions:

Do you have any comments on the wording or coding of these individual questions?

- 002: Q9 will be essentially uninterpretable for 'Christian', being chosen by many white residents as the default, who would actually express no religious faith or activity in lifestyle. Q11 Too vague - too wide a variety of interpretation.
- 003: I10 and I13 difficult to be clear what 'substantial' might encompass.
- 006: QI8 - I would very much like to see Irish included in list of ethnic groups. QI9 - I'd like to see Christian replaced by a range, especially Catholic/C of E, C of S etc.
- 007: I8 'Irish' is missing.
- 010: I8 Surely we should have Black British and Asian British by now for 2nd/3rd generation. I9 Why Christian for England, Wales, Scotland but denominations for N.Ireland. Cannot the EWS question at least divide into Anglican/RC/Free church/Other?
- 011: Particularly like QI10 - its high time 'Carers' were properly identifiable across the country. I13 More detail might be useful - i.e. who's cared for, broad disability category etc.
- 012: L1 Language questions need to distinguish language use and language ability. See Canadian Census 1990 for 'best practice!'
- 013: Any chance of a full matrix for relationship i.e. relationship to all household members. I8 Irish?
- 016: I can't help thinking that responses to I11 and I12 will be strongly correlated. A question about use of health service e.g. consulting GP in the last 3 months might give more specific information.
- 017: It would be useful to gender the relationship in the household e.g. grandfather, grandmother as alternatives. This would allow much better information on household structure e.g. matrifocal trends, matrilineal structures, and fit better minority extended family patterns and naming systems.
- 018: I would like to see some questions on pollution-related illness, say, I11a: Do you have any allergies? Yes; No. I11b: - Do you suffer from hay-fever/ asthma attacks/ enema/ respiratory-illness?
- 020: L1 Languages other than English spoken in the home. I8 Irish as ethnic group.
- 023: I10 & I13 Meaning of 'substantial'? I9 A waste of time as it is an amplification of I8.
- 027: I would drop religion outside Northern Ireland. I11 The general health question is really too vague. Would support the carers questions to go alongside long-term illness questions. I14: A simple write-in question would be shorter but not so accurate. I15: Extend this question to under one year of age by regarding either place of birth or mother's location one year age.
- 028: I5 What about those in part-time education? Will this be covered anywhere? Important for life time learning.
- 032: This is a hotch potch of cultural/colour terms. The policies of specific culture confuse it all. Do we want to know 'colour', cultural affinity, and all of these. Personally - I would prefer both - collected as a matrix, e.g. Black/Caribbean/fundamental Christian, etc.
- 034: I8 Why or why not include Irish which is the largest minority ethnic group in the UK? I9 Important new question. I10+I13 Essential new questions. I11 Again an essential new question. I12 I am glad that 'disability' has replaced 'handicapped' in the wording.

- 035: I very much would like Q19 to be included. This is essential to my research and we haven't had census question since 1851. Essential for comparison with other countries. A break down between Protestant and Catholic and other Christian (orthodox) would be helpful.
- 039: Why can't the proposed Northern Ireland question also be used in the rest of the UK? It may shed some further light.
- 040: I10. Good question. I11 Too vague.
- 043: Better to ask number of days of illness (unable to work or away out normal duties).
- 044: I7 Odd to call N.Ireland a country (slightly odd to call England, Scotland, and Wales countries if for 'other' you want something like 'Spain' not 'Catalonia'. I9 Why different for N Ireland? There are atheists (who will be offended), and Jewish, Muslim, etc., who have as much right to be listed there (and politically ludicrous to have N.Ireland different on this). I11 Meaningless, so useless without some explanation e.g. for your age compared to previous years, etc.
- 045: I14 is far too complex.
- 047: What guidance will be given on who is to be 'person number 1' - there shouldn't be a presumption of oldest, or which gender, but the form-filler should know this person is deemed the 'Head of Household'. I9 Why not separate 'Christian' into 'RC/C of E or Non-Conformist/other Christian'?
- 049: I13 is an important new addition but some concerns as to whether 'Yes, from someone in my household' is not too broad, likely to get a positive check from regarding domestic work/assistance - would be more useful to deprive in terms of childcare or care of sick dependent.
- 050: I19 only 2 categories needed: (1) never had a paid job; (2) write in number of years.
- 051: I29 The SWS appear to be very noisy, which suggests this question was poorly answered in 1991 (although I haven't read the quality assurance report, so I do not know what that has to say about the matter). The fact that this question has the same structure as in 1991 is therefore of some concern. Example of problems in 1991: For mode of transport to work, between wards in Leeds, there are a small number of non-zero entries for 'Underground'. This is presumably a confusion of over the term 'Metro' and as such is a local problem. Should people using local trains in West Yorks. be counted as using BR or as 'Underground, tube, metro'? If this is a problem then should it have been spotted in the edit stage? Is it a problem of wording? There are also problems with rather implausible results, such as the man who cycles from the New Forest to Kirklees, or the woman who walks from Worthing to Leeds. Presumably these, if genuine, are problems with differences between weekend and weekday residences. It is difficult to see how to address this issue without a major restructuring of the Census to allow for the small number of people who have two or more regularly used homes.
- 053: It would be helpful to separate schoolchildren status - as dependants - from students' status which may be dependent and also have some disposable income.
- 057: I5 - It would be useful to have information on people on part-time courses (e.g. day or block release) as well. I11 - Categories are rather nebulous - would it be better to ask something more specific (e.g. major illnesses or hospitalisation over previous year)?
- 058: T10 - Substantial - how defined? I11 - Good - how determined? I13 - Substantial - how defined?
- 059: Should give opportunity for recording of same-sex relationships, for example, 'living as a couple, including same-sex partners'. If not, then as in 1991, gay couples will frequently state 'not related', there will still be no data on this important issue for the study of urban social geography.
- 060: I10 - May need to define personal help. I11 - Need more categories to accord with similar surveys e.g. excellent, very good, good, fair, poor.

- 071: I14 - I think taught postgraduate degree and Ph.D. are sufficiently distinct to use separate wordings. I14, Is it too late to consider including asking if the qualification was obtained in the UK or not?
- 076: Should include Irish.
- 078: I10 - Could refer to the provision of childcare in which case no disability may be inferred. I15 - The wording will have to be much clearer. I17/I18 The phrase "last week" will make the questions difficult.
- 079: I9 - Poor wording 'Religious Group' - sounds like the minority! How about 'Do you consider you belong to a religion - or follow a religion'.
- 080: Should have an 'Irish' category.
- 081: I9/L1 ONS have not ruled out a question on mother tongue or language of the home: 'What is the main language you use at home: English; Other (write in).' Such a question would allow important social studies, add richness to the ethnic group codes, and identify areas likely to need extra support in early years education and translation of official materials (including the 2011 census!). A language question is of higher priority than religion which does not so directly indicate important service needs.
- 083: I10 & I13 - 'Substantial unpaid personal help' - is too vague - change to - more than 10 (or 20) hours per week. I11 - Add 'Very good'/or 'Excellent'.
- 084: I9 - Against this question on principle.
- 086: I10 - Needs some guidance on meaning of 'substantial'. Needs to include 'neighbour'. Need to include problems due to old age. I13 - Again 'substantial' needs defining.
- 088: I8 - Category for Arabs? I19 - Granularity of one year seems rather coarse - especially if you have just been made unemployed.
- 091: Very limited use of 'mixed ethnic group'. Religion question provides added value to ethnicity question. Welcome questions on unpaid help. Welcome expansion of education question.
- 094: Why does NI question exclude black-other?
- 095: I9 - 'Belong': there may be some confusion over, normally belonging or actively participating/attending religious meeting.
- 098: I am pleased with new question on 'caring'.
- 102: I11 - How do you judge these categories? - Need some guidelines! Hard to judge over 12 months. Better to ask: Have you been hospitalised? Have you had a major operation? Have you been confined to bed? - i.e. some more specific questions.
- 103: These questions are vague and unclear. 'Substantial' will be variously interpreted; the lack of specification of a time period is a disadvantage; the terms 'friend or relative' may or may not be thought to include neighbours or children.
- 105: I10 and I12 - What does 'substantial' mean? Vague.
- 111: I8 - 'mixed' - should be something like "parents from different ethnic groups". What about an Irish category? Black - other should be altered to enable people to positively choose 'Black - British'. With increasing numbers of SE Asians in UK, this might also be considered as an additional category. I13 - Question too vague: How about asking if a person has been hospitalised or had continuing medical treatment during the year?

- 112: Difficult to suggest improvement but I14 looks very forbidding. I24 - including regular overtime? I29 - Is there room for variability? Should 'yesterday' be specified. Could there be change - 10 yrs is a long time to get overall view of significant changes to the modal split.
- 116: I10 and I13: too similar or overlapping - will confuse? Is Part-time postgrad/vocational students?
- 120: I11 - "your health for someone of your age" has always seemed most useful to me, particularly if the question is combined with I12.
- 127: I8: It is essential that the category 'Irish' be added to the list of ethnic groups in question I8. The case for this is made in the Commission for Racial Equality's research report 'Discrimination and the Irish Community in Britain'. At present the 'Country of birth' question provides specific data on Irish Migrants but not the ethnic group as a whole. The CRE recommend a form of wording to include the Irish.
- 128: QI30 should have a section related to informal care - not just for child care.
- 130: QI18 Needs an Irish category - 1991 birthplace data showed similar levels of disadvantage to included 'Ethnic' groups - these are not addressed for the Irish community because they are unrecognised (largely due to Census categorisation). QI19 Needs to distinguish between Christian categories - at least Catholic and Protestant.
- 131: QI8 does not refer to ethnic group but national group. Also should include 'Caribbean' and specifically mention "Irish" "Arab" and not leave these simply as 'white' or 'other'. Finally - why should Northern Ireland exclude 'black' other' - since this may be precisely how a black person from there would describe themselves.
- 132: I am not a 'health grading' expert but Q 111 about general health seems to be rather too vague and general to be useful in assessing the important links between health and employment status. Maybe the health professionals have a better tested way of expressing it.
- 133: Relation with household: include something for foster child/foster parent? Q119 I think it would be sensible to include practising/non-practising for each category, otherwise don't bother with the question at all. Q114 potential for confusion between the categories.
- 134: If address is changed, it would be useful to know if the move was only housing related or job and housing related. This would facilitate much improved migration models.
- 136: Educational qualifications: seems very complicated - potential problems filling this in? Are you registered disabled would be a very useful question for us.
- 137: The initial question on relationship within household should consider changing the term 'living together as a couple' in some way to highlight the non-recognition of marriage by some such couples.
- 138: More detail on disability? Rather than being included in Limiting Long Term Illness?
- 139: I10 Who defines 'substantial'? I13 How does the question work when there is more than one carer? I30 could be misinterpreted as only applying to those who had work last week. Put 'in addition to any paid work' at the end of the question?

1.3 Wording/coding

Questionnaire instructions:

Do you have any comments on the wording or coding of these household questions?

Do you have any comments on the wording or coding of these individual questions?

- 002: Q14 is a mess, and will be left unanswered by many - but reflects chaotic educational policy rather than a deficiency in census design - too complex to work.
- 003: I17 interest in further breakdown of part-time (30 hours or less per week) category. I19 would prefer: 0-3 months, 3-6 months, 6-12 months write in number of years, never (important for policy to identify 6-month threshold). I28 useful to identify more than one usual workplace. I29 would be interesting, but not perhaps feasible, to identify different modes for different parts of week, and different combinations of modes.
- 009: I21 and I22 - Are there plans to ask these questions or the unemployed? i.e. 'What is/was your job/last job?' etc.
- 010: I14 seems highly fractionated - 4 broad groups would seem sufficient.
- 018: I would like to see some more questions on 'car accessibility', e.g. I29a: Reg. of vehicle used to travel work. Firm: employer; Firm: employer hired; Firm: h/h hired; Firm: not hired, h/h member; other person: borrowed; other person: no details. I29b: Driving licence. No; Full car/m'cycle; Full car only; Full m'cycle only; Full moped; Provisional car m'cycle; Provisional car; Provisional others. I29 c: Do you have day time access to a household vehicle? Yes; No. I29 d: Overall travelling time to work: less than 3 mins; 3-8 mins; 9-15; 16-30; 31-45; 46-60; 1hr -1.5hrs; 1.5-2hrs; 2-2.5hrs; 2.5-3hrs; over 3hrs.
- 019: I30 Presumably 'looking after the home or family' is intended to mean as a full time daytime activity. Many people would consider that they do this anyway as part of everyday life in addition to 'work'!!!
- 020: I30 Another job, other than the main job. I29 Distinguish between working At home and From home.
- 025: I recall I29 as being difficult in 1991 if one travelled in avariety of ways.
- 027: I27: I think this assumes knowledge that only some people will have. I28: Improve this question by asking respondents for postcode and as much detail as possible e.g. University of Leeds, Woodhouse Lane, Leeds LS2 9JT.
- 028: Again what about something on training/education currently being on a part time basis. Part time further education and high education needed here as a separate category.
- 031: I27 'at the place you work for' needs rephrasing.
- 034: I14 This question will be of much, much greater value than the old wording. I18 Important new question. I19 Important new question.
- 035: I14 Will you be able to code if they are universally as a particular level of the student/pupil? Very useful question.
- 038: I27 'at the place you work for...' is an odd wording. I do not believe that many respondents will be able to give an accurate answer to this question anyway.
- 039: In I29 how are journeys that are multi-modal to be dealt with?
- 044: I21 Why are all these services? Could bias results. I23 People supervise 1-2 others without considering themselves on being treated as managers. Needs rephrasing. I26 Answers likely to be too diffuse to be useful for occupational use. I29 This must allow more than one answer. It should allow indicating longs

- + distance where several modes used. I30 Doing ... activities: awful phrasing. Why only government training? I would be interested in any training.
- 047: Does the form-filler only enter highest qualification? Is there a box to put in, for example, the subject of the degree? I23 should differentiate between a supervisor/manager in a small shop (say) or of a small group of workers, and someone with responsibility for (say) at least 10 other employees.
- 050: I29 Metro or light train needs to be a separate category (e.g. for Tyne-Wear metro and Manchester's Metrolink. Also add 'tram' (e.g. for Sheffield's Supertram), separate from 'underground/tube' and from 'train'.
- 055: I14: Find detail and question is much needed - current question not much use for many purposes
- 056: I14 category (2) = illustrate with example? How does (13) differ from (15)? I22 is probably going to be very difficult to classify. I24 Breakdown. Normal and overtime. I27 Category 500+ is too big. i.e. use 500-1000; 1000-2000; 3000-5000; 5000+.
- 057: I14 - Great improvement on 1991! I19 - For the purpose of labour market analysis, it would be better to have more ranges, with the lowest being at least '12 months or less'. I30 - Again, need to cover those in part-time education.
- 058: I22 - How defined? How many? I28 - Postcode?
- 059: Have Ph.D. as separate category - there is really no similarity between the PGCE, for example, - a 3-4 year Ph.D. Aggregate data would be misleading.
- 065: I28 The category 'mainly at/from home' should exclude people with live-in jobs in hotels etc., and people who hold their accommodation with a job, shop, farm, etc. i.e. people living at work. I21-I25 - All these questions should be asked for last job in last 10 years as well as current main jobs.
- 068: I would like to see the band for 25-4999 people employed by the organisation (I27) split into one or two further categories. As it stands, it is too wide.
- 071: I23 - Some jobs include demanding supervision of other people who are not employees.
- 078: I30 - May include voluntary work either for individuals or for a charitable organisation.
- 080: I19 - Ten year period too broad.
- 081: I17 - Those who worked casually, paid direct in cash, or piece-work at home, would find it easy to say no to this question, as 'employee' suggests an official status, but should be included. The question would be better 'Last week, did you do any paid work? Self-employed/freelance; Employee ft(); Employee pt (); Other paid work. I28 - Delete 'or from' in second category. It can only mean no fixed place (e.g. a builder) and such people can tick the first category. That will leave 'Mainly at home' to allow for the first time an analysis of those who do work at home.
- 083: I19 - Need number of years since had a paid job for all respondents. I18 - No. of paid jobs as 'self-employed/freelance' may be misunderstood - or number of self-employed contracts.
- 084: I14 - Looks excessive. Just how reliable were the answers in the testing? I23 - This seems to be of little use really. Is it worth asking a question on this?
- 088: I28 - Could working at home be treated separately to working from home. It would be useful to compare these two groups. I27 - What about home workers working for a large organisation? How would they answer this question? I28 & I29 - For main job? Needs to be made clear.
- 093: I19: The 10 years cut-off seems very long. I22 - Seems too imprecise for worthwhile response.

- 103: I29 - Will people have opportunity for multiple codes? I30 - Could add the code - 'doing voluntary/unpaid work'.
- 111: I15 - Prompt respondents for their former postcode? Five - year migration question? What about a question category on "Helping in Family Business"? (maybe unpaid). I24 - Expand to take account of 'flexible' employment practices, e.g. shift work, etc. I29 - Specify major usual mode of travel. Add time taken for Journey. Need to take account of increasing diversity of work journeys. How to deal with people self-employed but not working at home - e.g. tradesmen. How about an explicitly 'Teleworking' category?
- 116: I14: Pre-codes too detailed - might produce biased response. I26: Respondents will be an unreliable source for this information? I29: Mode used for longest part of journey. I30: + on holiday, temporarily sick.
- 117: *Adoption*: Relationships *Separate* son or daughter from *Adopted* child born to another parent - adopted by parent's husband or wife. It is important to gain reliable information about numbers of adopted children whether these are 'stranger' adoption or 'relative' adoption. It is important to gain separate information on adoption which is not available from any other source. Adopted by both parents/single parent - non relative. Child of one parent adopted by husband/wife. Adopted by relative. If adopted/fostered - age when joined family.
- 117: The *ethnic question* was not satisfactory last time. Particular confusion between "black" and "mixed". The 'mixed ethnic group' was not satisfactory last time. Add 3 categories - One black parent and one white parent - one South Asian parent and one white parent - other mixed ethnic groups (please write in). Someone of mixed parentage e.g. African-Caribbean father, white mother, might have put Black-Caribbean. Another person of same ethnic background would have put "mixed ethnic background". It would be clearer if, instead of "Black African" it read: Both parents Black-African- Both parents Black Caribbean - Both parents of other Black ethnic groups - Both parents Indian etc. Mixed ethnic origin parents of different minority ethnic groups (please write in). One Black African parent, one white parent - One Black Caribbean parent, one white parent - One Asian parent, one white parent (write in Pakistan, Indian, Chinese, Bangladeshi or other).
- 117: *Irregular* employment is a particular source of stress. How many change of main job did you have last year (please write in number of changes of main job).
- 123: Language spoken at home (mother tongue).
- 127: I think a question about religion should be included throughout the UK for the following reasons: a) There is currently no way of differentiating migrants from Northern Ireland or their descendants in terms of coming from a Protestant or Catholic background. This is regrettable as there is some evidence that they have different experiences settling in Britain (see Hornsby-Smith & Dale, BJS 1998). b) The CRE report suggests that there continues to be differentiated socio-economic experiences for Catholics compared with Protestants in Scotland. c) I support the case of some Muslim groups for the inclusion of a question on religion. d) For all these reasons 'Christian' as a category should be sub-divided as it is in the NI Census - otherwise it will produce a Christian/other binary just as the ethnic question regenerates a misleading white/black binary.
- 128: As stated earlier there is a need for an Irish category in the ethnicity question. CRE reports high incidence of discrimination and disadvantage in housing, employment, health and criminal justice, despite data which grossly under estimates. The Irish are a distinct cultural group despite their fair skins and debates constructed around a black/white binary fail to recognise the special situation of the Irish in Britain. The Irish are an older population with distinct health and community care needs which is not tackled strategically.
- 129: I would like to see 'Irish' included in the categories for ethnicity (as recommended by the CRE). My research concerns the Irish community in Britain as well as other migrant groups. Religion as a category should be included (Christian broken down by R.C. etc.) of particular interest in relation to those of N. Irish origin.

131: Q130 - in main body of question, the word "paid" should be inserted between those if 'any' and 'work' - and this should be applied to all similar questions. After - "looking after the home..." should insert "has this been the usual pattern for the last 3 (6) months?"

1.4 Income question

Questionnaire instructions:

Do you have any comments on the wording or coding of this individual question?

- 002: Should specify that the answer can be given per week *or* per year to avoid two answers which may be incompatible, hence uninterpretable.
- 003: There is difficulty in allocating some state benefits, interest, etc. between individuals in a household. A question of earned income would be easier to answer at the individual level.
- 007: Perfect -- but ask for over £100k/year.
- 009: If possible a tick box on 'Do you receive one of the following: Benefit/Income Support/Lone Parent Allowance/Disability Allowance?' A question on Savings/wealth? (similar to Income Bands?)
- 010: Another group for £25-30,000; £30-50,000; over £50,000
- 011: Excellent idea to include an income question, but is the upper category of £25,000 or more really going to tell us anything? More normal families earn more than that and how do you distinguish these any degree of wealth.
- 014: We need to include more boxes at the top and to avoid too much aggregation there (national average GDP/capita is almost £20,000 too many people will fall in the open-ended top category).
- 018: Is it possible to expand the number of income bands?
- | Category | Per year (approx.) |
|----------|--------------------|
| (6) | £15000 - 19999 |
| (7) | £20000 - 24999 |
| (8) | £25000 - 29999 |
| (9) | £30000 - 34999 |
| (10) | £35000 - 39999 |
| (11) | £40000 - 49999 |
| (12) | £50000 or more |
- These will provide vital information on what can they (high-income group) buy/own?
- 019: Perhaps some reference to time should be included. i.e. Do you mean in the year ending on census day? The previous calendar year? Previous financial year? What about 'windfall' payments—should they be included? Inheritance? Lottery winnings?
- 020: The upper bracket is too low. In anticipation of data comparability with 2001 there should be at least another above £25,000.
- 023: We need more upper categories - over £50k and over £100k are essential. The category £25+ includes people in quite modest professional jobs and the richest persons in the UK.
- 025: Is one mention of the word 'gross' an adequate way of specifying pre-deduction income? Is it clear that 'your' applies to the individual? £25,000 seems rather low as a lower band for the highest category.
- 027: I think the final category probably needs to be split up further. It might be sensible to use the income bands that define deciles of the income distribution.
- 031: More bands: £25,000 - 49,999; £50,000 or more.
- 033: May be worth putting in monthly equivalents

- 034: This is far and away the most important new question. It will make the census far more valuable. There are two major problems with the wording. First the top band is too low (£25,000+) and it should at very least be set at the top rate of tax income (currently £26,100 - but will be higher in 2001). Second, net income is of more use than gross - Britain is alone in Europe in asking about gross income - see 1991 EUROSTAT Harmonisation Workshop. Surely every other statistical office in Europe can't be wrong?
- 035: Do wives put in husbands' allowance also?
- 038: The threshold for category 7 is too low. The question will lack distinction for high incomes. Suggest £30,000 as the threshold for the top category. In coding this question I would suggest that a new variable of 'total household income' should be created from the individual answers for any given household.
- 040: May be useful to add monthly figures too.
- 041: It needs to be clear pre or after tax/NI/pension etc.?
- 042: Need to take banding higher than £25K.
- 043: Better to fix income boundary in relation to minimum wage or level for receiving supplementary benefits.
- 044: Very low top point for fiscal/distribution work. Need to specify if includes capital transfers (e.g. B.S. Windfalls). How many people will know (or calculate) gross interest, etc.? Does this include parental distributions to students?
- 045: Is (7) too low a top group? (average earnings are already in category 6).
- 047: A few more categories, breaking down the higher earners, would be valuable (e.g. £25k-<£35k; £35k-<£45k; £45k+)
- 048: I31 Needs further categories above £25k.
- 050: (1) After 'gross' add 'that is, before any tax is deducted'. (2) Add an extra income category e.g. £25,000 to £40,000; over 40,000 to better identify higher rate tax payers and allow better comparability over time if question used in late Census.
- 051: I31 A question on income would be very desirable, although the concerns surrounding it are well known. The question as posed seems to have an upper limit which is rather low.
- 056: A question reaching to the sources would be useful.
- 057: I would like to see 2-3 more ranges included: e.g. £15,000 - 19,999; £20,000 - 24,999; £25,000 - £99,999; £100,000 or more.
- 058: (7) £25,000 to 35,000; (8) £35,000 to 50,000; (9) £50,000 or more.
- 059: Need a 8th category, £40,000 or more, for the study of wealthy city populations, £25k is too low a maximum.
- 064: Top end range not sufficiently high.
- 067: Research using banded earning data is difficult because of the different size of the class intervals. It would be more useful if the respondents specified their total gross income as a single amount.
- 068: Has to be included in the 2001 Census but current question needs to be changed. The question should be more specific describing the elements of total income in more detail (perhaps in an appendix). The bands are too wide, particularly at the upper end of the income distribution. Just 2 categories for >£15,000 when average earnings from employment were around £18,000 in 1996 (NES). A lot of people will be classified

in the over £25,000 band which makes meaningful analysis almost impossible. The top band should be set at a much higher level.

- 071: I think £40k or more would be useful.
- 076: Top of range far too low. Should be £25-34,999; £35,000-44,999; £45,000+
- 077: Needs to be clearer whether inclusive or exclusive of income Tax, NI, Pension etc.
- 079: The category Rent could better be called Rent Allowance or Housing Benefit/Rent Rebate. But strictly these housing allowances could be argued to be excluded from Total Gross income on the grounds of being housing subsidy as opposed to part of Income Support.
- 083: Add year income before tax and deduction for N.I.
- 084: No. Against I31 - I don't like it and suspect it wouldn't be answered faithfully anyway.
- 091: Upper band on income too broad.
- 093: Perhaps needs 1 or 2 extra categories of top level.
- 094: Do people understand what the difference between gross income and net income?
- 097: I would prefer a more exact measure. £25,000 or more is too low for a top category.
- 098: I think at least one further high income category would be warranted (e.g. £50,000 or more) - and will it be clear whether it is individual or household income?
- 100: It is clear whether respondent should include income to cover housing costs e.g. mortgage interest payments in income support, housing benefit.
- 102: Categories of income only up to £25k and >£25k. Helpful to have a fewer breakdown at higher incomes: £25-£50k; £50-£100k; >£100k.
- 110: Need for interviewers to make sure that respondents understand the term gross income from all sources.
- 111: Upper bound of £25,000 far too low, especially in London/SE. Difficult to know how to deal with benefits - presumably household benefits should only be answered by household head. There might be double-counting. Also problem of confidentiality - not all household members (e.g. lodgers) might want form-filler to know their income.
- 112: Make more explicit differentiation between total gross household income and personal income. Why stop of £25,000? - graded to £50,000?
- 116: Before tax and deductions?
- 119: Although many might argue that this is one of the most important questions, it might also turn out to be the most misleading. Maybe a 'per year' category should also be included as a means of double-checking. The band should certainly be extended beyond £25,000 per year and additional categories should be added.
- 120: This question is of interest but highly problematic (a) degrading accuracy of response (b) affecting overall response to the census.
- 124: Prelude BEFORE TAX, insert before 'falls'. Highly desirable to have more than seven categories in view of complex and non linear relationships of income to health. See E Cob and Davey Smith The nature of the relationship income to health (submitted Social Sciences and Medicine). I have made this point in

separate correspondence with Angela Dale in respect of non-occupational indices of deprivation and the crucial role of household equivalent income in such a measure. (Details available on request).

- 132: To me the obvious request is for a further income band of, say, £40,000 or more. When it comes to issues of polarisation in large cities, particularly London, an upper limit of £25,000 has only a limited value.
- 136: £25,000 seems a rather low threshold for a maximum category.
- 139: Make it clear that intra household payments are not counted as sources, i.e. say “Any other sources from outside the household”. Perhaps it could actually say, this does not include allowances or pocket money paid by another member of the household.

1.5 Other questions

Questionnaire instructions:

Are there any other questions which you feel must be included in the 2001 Census?

- 007: What is your total gross wealth from all source: saving, house equity, belongings (insurance value). It isn't difficult to ask this question.
- 009: A question on Savings or wealth e.g.: Saving ranges £0-£1000/ £1001-£10000 etc. A question on registration letter or year of car was built. A question on receipt or benefits.
- 010: I8 Having worked with the frustrations of the ethnic categories of 1991 and listened to the frustrations of many of the non-white inhabitants of my city (Birmingham), I approve the introduction of the 'Mixed' category but deplore the failure to introduce categories for British non-white ethnicities. Black British and British Asian are essential for those 2nd/3rd generation who are British first and Black/Asian second. The other hidden group in 1991 (and seemingly again in 2001) are the 'African Asian' i.e. those south Asians who moved to Britain from East Africa in the 1960s and 1970s. There is a significant 'Arab' population in some cities which may well be nearly as big as the 'Chinese'. I9 Most of my work with the census has been for church organisations. Since we are at least to get a religious question (which will be important in linking with ethnicity) please, please don't waste the opportunity and leave 'Christian' unfractionated for England/Wales and Scotland. A fourfold division into Anglican, Roman Catholic, Free Church (Baptist, Methodist, URC), and Other (please write in) would be immensely useful in my work both academically on urban culture and ethnicity and in enabling religious groups to make use of the census.
- 014: I think we should distinguish between labour earnings and income from other sources (transfers a profits/rents).
- 027: (1) The Address One Year Ago question must capture under one migrants. This information is needed in any application of migration data to population estimation or population forecasting. If it is not extended, can the Census Office adopt the convention of reporting the under ones by the mother's address one year ago. The Census Office should consult with the Canadian and Australia Census Office where this problem arises and has been solved. (2) Adding to I30 Economic Status one year ago would be very valuable in tracking status changes along with migration.
- 028: With the emphasis on life time learning more information is needed on current training/education of those not in full time education but role of employer/government support. Increasingly people will acquire skills in this way.
- 033: I14: This seems important given solidity and growth of post 16 education.
- 034: A question on whether the household had a telephone would be very useful, as (1) a poverty indication; (2) a indication of social isolation.
- 037: If you ask QI29 'Travel to work' then 'Travel to place of study' should be included.
- 044: Many of the questions are awkward or unclear. They badly need work by the Plain English Campaign. For several, I would not be sufficiently confident that respondents had understood them to use the results. For some, e.g. the use of 'country' to mean any of Wales/England and Wales/UK. I have no confidence in the compilers' understanding.
- 047: Main job one year ago: the value of migration data, and the understanding of the labour market and the experience of young people (e.g. the benefits of training and education) would be hugely improved by knowing the jobs people were doing 12 months ago (the date for which they say where they were living).
- 048: Employment of servants (cleaners, nannies, as pair, etc.) within household. (observation suggests this is a rapidly growing phenomenon, but there are no reliable data on its present size).

- 050: Given that work trips account for only 20% of all travel, it would be highly desirable for transport/land planning to have a question on, say, food shopping trips.
- 056: Training in the last 12 months. Type of training received e.g. off-the-job; in-company; in-the-job.
- 059: I think it is vital that the household question (see pp.6) makes a clear that same-sex relationship can be recorded. We have no reliable data on this at present and it would be very valuable for the study of the large UK cities. Given the change in gay lifestyle, this is more relevant than earlier centuries.
- 064: Extended marital status information, e.g. number of times married.
- 068: Use of English language and other spoken at home/work. This variable will give an important insight to the cultural differences which exist within Britain. Also, ability in the English language has an important influence on a number of outcome variables e.g. employment and qualifications.
- 071: (1) Country of obtaining any degree level qualification. If one is trying to understand the mechanism of obtaining a degree in the UK, e.g. effects of ethnicity, gender, then it is vital to know if the degree obtained was obtained in the UK. The circumstances of an African person, for example, obtaining a degree in the UK are virtually different from obtaining that degree in their own country. (2) Within qualifications ask whether obtained as a 'mature student'. The probable circumstances of mature and non-mature status would be expected to be quite different. The present government develops education is its top priority. Also, in recent years university participation has risen to around 30%. If we wish to understand the process at work in this field then more detail is useful: especially where they get their degree (UK or not) and the process (mature or not).
- 077: (1) Category 'Lone Parents' to be widened when data released - definition appears to exclude parents with any child over age of 15? (2) More detailed release required of data on educational achievement - especially poor or no educational qualifications.
- 078: The public's support for charity has only been measured using very small samples. Donations of all sorts (cf. The Individual Giving and Volunteering Survey published by CAF) ought to be measured in a much more rigorous way. A question asking about the cause supported would be beneficial. I would suggest a question on the amount given to charity in the previous week (perhaps within certain categories) and both a checklist of categories of cause (e.g. children, cancer charities etc. - tick all that apply) and an open box to list the names of charities supported. There is, quite simply, no other way that charities will gain an objective view of the nature of UK donors. There is also no likelihood of charities ever having the resources to fund such research.
- 081: Ability in English. A question on language and an indicator of ability in English is essential for reasons given on page 6. However, ability in English is probably too difficult to ask in a census. 'What is the main language you use at home? (English; Other/write in) could be mainly machine coded and users including myself would be happy to take the number of people whose main home language was not English as an indication of the number of people whose English was not fluent, for analysis and prediction of service demand.
- 085: Economic position one year ago or (at minimum) student or not one year ago.
- 088: The travel to work question could be changed to enable it to capture information on travel to place of study for students/school children. Time taken to travel (on average) in minutes. It would be interesting to look a travel times for particular sub-groups (e.g. female part-time workers with children).
- 090: 'What is the main language spoken in your home' 'What additional languages are spoken in your home?' Answers to these questions will provide essential information for planning and providing services such as education, speech therapy, advocacy in health matters.
- 093: (1) 5 year migration question. (2) Address of graduates in year after graduation (i.e. graduates in 2000) - where studied, home or elsewhere.

- 097: English ability. Also an earnings question to supplement income question, i.e. remuneration from paid employment.
- 110: Current level of savings.
- 111: Language: There is a clear and consistent demand for information on the languages spoken by the population, especially (but not exclusively) South Asian languages. Needs for this information: 1. Planning education provision - native language support for children. 2. Planning Social Services provision/Health Service provision for older people and women. 3. Determining the need for information provision in minority languages. 4. Enabling Private sector organisations to identify markets for reading matter, products and media (e.g. local radio) in minority languages more effectively.
- 116: Increasing number of people in several age groups have complex residential patterns involving more than one address. The 'fluidity' likely to increase. Important that population census begins to trace the prevalence and geographical patterns of the phenomenon.
- 119: 'Ability in English' would be a very useful question as mobility from the EU countries is expected to increase, but it should be carefully worded.
- 137: A more particular question that allows identification of the number of families within a household would be helpful but I do recognise that this can be worked out from the main relationship within household question.
- 139: Relationship grid of all household members. It is difficult to find lone parents in complex households without this information. LS has difficulty matching non-members across censuses in these cases too, for such purposes as looking for step-relationships.

PART 2:

PROPOSED CHANGES IN CONCEPTS FOR THE 2001 CENSUS

2.1 Population base

Questionnaire instructions:

Do you have any comments on the Population Base proposal?

- 001: Comparability between 1991 and 2001 is essential.
- 002: Regardless of the base chosen, backward comparability of key counts is essential.
- 007: Missing a million people is more important!
- 008: Continuity is vital - even if the proposed change were desirable in itself, I would oppose it if it threatened continuity.
- 016: The proposed change would facilitate comparisons with other register data e.g. those held by FHSAs.
- 019: My research is based on small numbers of childhood cancers over long periods of time (from 1962 onwards). Hence comparability between census is essential.
- 027: This is definitely the right way to go, but as much help on temporal comparisons should be provided as possible. A 1991 Population Base figure for all 2001 Census outputs will be essential.
- 029: The estimates of population change over time is essential to many users of census. This alteration to the population base will undermine its use.
- 033: Much better to do as vacation address plus flow - keeping continuity with 1991's.
- 043: Strong support for the change.
- 047: The population base is clearly an issue for which there is not a single correct answer; the Census Offices seem over the last 30 years to have got the single incorrect answer i.e. changing it every time! What is needed is comparability through time - if that is impossible (as it may now be), then comparability across the EU (or as much of it as possible).
- 048: The rising cost of studentship will force a growing population of students back into the parental home.
- 049: In terms of use made of the LS it is desirable to maintain some close means of creating equivalence in household even population bases. This could be achieved through the SARs?
- 051: The changes suggested in the population base are sensible; what is essential is that very clear and complete information is provided on the differences between the 2001 base and the 1991 base and other past bases. As well as basic counts using other bases in Table SAS 1, there should ideally be a user guide (or equivalent) which would provide pseudo-code algorithm demonstrating which counts should be added to or subtracted from the bases in order to allow reasonable comparisons, and specific notes about which SAS tables can easily be compared and which cannot. Preferably this should be reasonably exhaustive, i.e. the calculation generate a pseudo 1991 base, pseudo 1981 base etc. count from the 2001 data should be shown, together with calculations to convert 1991 etc. to the 2001 base. If these calculations could come with an advisory notice about how accurate they are likely to be, then so much the better. E.g. 'to convert 1991 data into the 2001 base, do a+b-c, this should give a close estimate' 'to convert 1981 data to 2001 base, do a+b-d, this will give a less good estimate, because d isn't really right but it's the best you can do'.

- 057: There seems to be no reason why two different population bases cannot be provided. To avoid any confusion, there could be an issue of the Census User Guide setting out their precise composition, and the appropriate applications in which they should be used.
- 059: It is obvious vital that small-area changes can be recorded, appropriate for the 1991-2001 period.
- 073: I see no reason why the base cannot be varied.
- 074: We require high-quality population figures for each year in the changes of Census data.
- 085: But how? At present, there is no question on the form which allows the 2001 population to be calculated on 2001 definition. Either this question needs indicating, or the 1991 Census has to be used to estimate the 1991 population on base of 2001 definition.
- 093: Generally supportive.
- 094: It will make longitudinal studies difficult! Students are not part of the long term population of an area but are important - can they be separate out and treated differently?
- 101: Essential that students identified in all tables.
- 102: It would be very helpful to have population base in line with one-year estimates but probably more important for research to have comparability with 91 - both population counts should be available.
- 106: Inevitable changes in GIS definition etc. will make comparisons very hard anyway!
- 111: Much of my work on the 2001 Census will involve analysis of 1991-2001 change. Any change to the Base complicates this and I don't think an additional population count would be adequate.
- 116: Why not tackle more comprehensively the recording of 2+ addresses. Could be done in two or three simple questions.
- 117: It is important to know about the 'connectedness' of students with families, long term family support may have an important impact on life choices.
- 136: Manchester has such a large student population that suggested base is a very useful one for us. However, it would be helpful if a comparison could be made with the 1991 Census somewhere in the tables, as in Table 1 in the 1991 output.
- 139: Being able to relate students to their families of origin increases the scope for linking them to the earlier circumstances of their families, and increases the possibility for intergenerational comparisons (e.g. social mobility) at the next census.

2.2 One number census

Questionnaire instructions:

Please add any advice on methods for producing a one number census.

- 001: Comparability between 1991 and 2001 is essential.
- 007: A one number census will redeem the Census Offices for errors after 1991.
- 016: Organising a national programme to assess the completeness and reliability of FHSA registers might be useful. This has been done in individual cases but confidentiality concerns, among others, seem to have prevented a more comprehensive evaluation.
- 017: The mechanisms for avoiding replicates should be very transparent.
- 019: In the past there have been so many alternative sets of estimates it has been difficult to decide which are best or most appropriate. It would also be very nice if all the estimates in the different tables were consistent. For example population by single years of age in one table (i.e. age 0,1,2,3,4) should be the same when summed as population 0-4 in any other table. And populations given in various tables should be the same.
- 024: Would it be that much trouble to distribute raw and adjusted figures?
- 027: (1) The key problem will be to determine how many people are missing in each area through statistical comparison of 2001 Census returns against alternative counts, to establish, as with EwC work, which are key indicators. (2) The missing people will need to be allocated to the smallest building brick (probably the unit postcode) but this will cause difficulties: How to invent imputed addresses will be an issue. So perhaps they should be allocated to output areas and assigned the OA grid reference. (3) Missing people will need to be added as both whole households and as extra persons in existing households. Simple random assignment of persons to households would probably not be sufficient. (4) The Census Office have given themselves an enormous challenge! (5) A fall back position of outputs without the added missing persons except for revised age/sex counts might be needed therefore.
- 034: I don't think it will be a problem in 2001 now the poll tax has gone. I don't believe that any of the proposed techniques (capture/recapture, etc.) can be of anything but marginal help. The best way to get one number is to make effort to enumerate everyone. Much effort is needed to enumerate the homeless!
- 041: I think any imputation using Capture-Recapture methods must be treated with extreme caution! Interaction between sources can very quickly alter (in either direction) counts of 'missing' persons.
- 044: Does this mean no publication of unadjusted data (or, same thing? of the method of adjustment)? Size and method of adjustment obviously essential, so what is meant by 'one number'? Sounds like slogan, not serious concept. For comparison with other countries (my interest) unadjusted counts are necessary.
- 047: An indication of the extent of imputation (or whatever) applied to the raw data in each small area is essential as part of the published data.
- 051: This is appealing in theory. It should be remembered that many problems in a variety of datasets produced from the 1991 Census were only discovered by users, who were able to put a much greater collective effort into the effort of error detection than OPCS could, or ONS will be able to in future. Perhaps ONS could do an initial release of data to interested parties (academics and other groups who are prepared to work with the data in the knowledge that some values may change) as part of a final testing phase; i.e. make sure that the initial release is referred to as test rather than as a release!
- 081: It is likely that non-response will again have characteristics that depend on age, sex, locality, and social or housing characteristics. Even if the validation survey is highly successful, such that users know the characteristics of non-response, they could not adjust census output themselves because they would require

the tabulations that interest them further disaggregated by all the variables associated with non-response. The one number census proposal, for an augmented data base of individual and household records is therefore the only approach which will satisfy the many applications which are significantly biased by census non-response.

- 085: Highly desirable to get close to the full population but do not let imputation and estimation possibilities reduce the attention and care given to getting back real Census form returns. This also requires a much fuller and more comprehensive post-enumeration survey, as well as a lot of preparatory work on rolled forward estimates.
- 086: It is important for users to know the methodology by which the 'one numbers' are derived - particularly in terms of households.
- 089: Although the caveats concerning comparability with previous years, the reliability of the method and assumptions made would be of considerable importance in analysis.
- 098: Use the national media to stress the confidentiality of the Census in the year or so before. Employ extra, especially - trained (and protected) enumerators for 'hard to enumerate' population.
- 102: Good idea, but must show estimation method for missing persons is robust!
- 116: I predict resistance from local authorities.
- 124: Information must be retained on whether the census responses is real or imputed and if the latter, on the nature of imputation used. This is essential in estimating the reliability of the figures produced. Missing records from 2001 can be carried over from duplicated cases. 1991 Census. Thought should be given to multiple imputation related to this (a la Rubin and Little) as an additional alternative to the 'number one census'.
- 130: I am concerned that the method used may be highly inaccurate for specific populations e.g. Irish homeless. No specific correction factor was published after 1991 for the Irish although other Ethnic groups/ages/locations were taken into account.
- 139: Keep imputed people out of micro data, or at least flag them. Hope it will not delay results too much.

2.3 Processing timetable

Questionnaire instructions:

Comments on the Processing and Publication Timetable.

- 002: Inevitably, users want it all. The planned versions idea seems reasonable, and 100% right first time virtually impossible!
- 007: You can do both-again, it is no longer difficult.
- 019: I see no advantage in processing parts of the country before others. However if the full datasets are to take a long time to compile it would be useful to have simple population counts as early as possible.
- 027: My view is that there will have to be a sequence of releases of data and that an organised programme would be better than continual crisis management under pressure.
- 034: 100% correct is impossible, post back (on its own) in the US does not achieve better than a 50% return rate. Early output with some errors is better than no early output.
- 044: (1) Here country means UK? Not England, etc., as in I7? If processors don't know what country means, how can respondents? (2) Is 'nation' the same as country? (3) Not only unrealistic, but seriously misleading. It must be adequately emphasised in publishing results that there is uncertainty, and errors or incompleteness will be revised. There is serious confusion in this section between public relations and method.
- 047: '100% correct' is a wildly unrealistic target on past records, so inordinate delays should not be tolerated in pursuit of this chimera.
- 051: The strategy of waiting until all returns have been collected is obviously an integral part of the one number census. It would also be very useful for the production of useful tables in the SAS (or 2001 equivalent) that have been compiled from interaction data.
- 057: It is difficult to answer these without information on their projected timetables. However, any view would be that the principles governing release, in order of priority, should be: (1) accuracy and comprehensiveness; (2) speed of provision.
- 064: The data has to be complete, correct, and globally available at the same release data.
- 073: It is a very long time since 1991 already.
- 081: I would like to see a contractual agreement to a timetable of quality checks and outputs which will encourage procedure to (a) get it correct first time and (b) correct errors quickly.
- 083: All data should be available for purchase on CD-ROM.
- 084: We should not need to wait for it all to be processed. If it is available, release it.
- 085: Timetable, OK, if it is a One-Number Census and as long as users are given chance to check data. Some basic preliminary counts would be useful if not to be a ONC.
- 091: Better to get it right than to get more quick results which are incorrect.
- 093: Producing a series of version releases seems a recipe for disaster.
- 094: The problem with Census data that by the time its published is already out of date - it will now be more out of date.

- 102: One preliminary report could be useful.
- 108: Details of additions to raw census counts required.
- 111: After every Census, the problem of the delay in getting results attracts considerable attention. Real World Census users need results quickly as well as accurate data. What will be the implications for SMS/SWS database? Will these appear even later than usual?
- 116: Must try to meet priorities of different categories of users.
- 119: I agree with the sceptics that 'one hundred percent census' is unrealistic and I would like to add that it is very unlikely that we see any results within one year especially with the post back method. Still, it looks as if this is the most sensible route to follow.
124. Depends on when complete data are available! 1 year, 2 years! 3 years!. Hopefully desirable to have data within 1 year for which there is some confidence. Suggests plan release within 1 year - if possible with complete data.
130. Preliminary data would allow ideas to be developed which could then be adjusted when the precise details became available.
- 132: I support the argument that, whilst the concept of a "100% accurate Census" on publication is a good one, experience suggests that its chances of delivery are less than 50:50. If it is proclaimed and then fails, the consequences will be damaging to the planning [of the next census].
- 136: Any delays in releasing 2001 Census data must be avoided although not at the expense of accuracy. The above suggestion [one number census] on paper seems OK but only if a speedy output can be guaranteed.
- 139: My interest is to see a faster LS link than last time!

PART 3:

PROPOSED CHANGES IN OUTPUTS FOR THE 2001 CENSUS AND USE OF THE 1991 CENSUS

3.1 Output areas

Questionnaire instructions:

Comments on output areas from the 2001 Census.

- 001: Essential to be able to link 1991 to 2001 areas.
- 002: 2001 EDs will be very different to 1991 EDs. If users have to cope with a new geography (inevitable) then make it a useful one - OAs built from postcodes and 1991 EDs/OAs would be great for comparative work, but useless for 2001 detail in areas of statutory boundary change or significant population redistribution.
- 007: You can't get 1991 EDs wrong very easily, so use them and subdivide them into OAs.
- 008: I need ward-level comparability, which all the options seem to provide.
- 009: Comparability between 2001 and 1991 Census is the essential element here. If OA's built up from postcodes are the option chosen, a look up table linking to 1991 EDs would be extremely useful.
- 016: As close a match as possible with postcodes is important. The ED/PC directory has been a considerable source of frustration for me. There are still (3rd release) lots of ridiculous errors in it - why hasn't consistency been checked more thoroughly with a GIS?
- 027: There is substantial support across all customer sectors for the Output Area proposal. Other needs can be met (approximately) by developing high quality look up tables. The academic community should have a programme for working with the Census Office on producing LUTs, and on using them to produce a library of area statistics for a range of small area geographies.
- 028: The worry within postcodes is great. Is the PAF file complete enough? It is an excellent approach but is it readable?
- 034: Postcodes are not useful for constructing social geographies. They do not divide socially homogeneous areas.
- 041: I really want to see postcodes aggregate exactly to output areas.
- 047: There is no point in making OAs literally match Byzantine boundaries such as those of parishes or National Parks - or even wards, which change frequently - because the data is only a 'best fit' measure (due to imputation etc.), and is 3+ years old (at best).
- 048: Change over time is an important aspect of health and deprivation analysis.
- 051: A lot of work needs to be carried out in the investigation of the nature of risk of disclosure. The current thresholds are useful in the sense that they are easy to explain and understand, but the specific numbers chosen have no theoretical backing. 2001 EDs should only be considered as a fallback position in the worst case. They are designed with collection in mind, and this will not necessarily be of use in analysis. OAs built from postcodes would seem to be the best choice, assuming good LUTs are available which provide an easy way of getting good estimates of other (past/contemporary) geographies. 1991 geographies would be useful for temporal comparisons, but would probably generate too many problems where new dwellings have been constructed etc.
- 054: Loss of exact electoral ward figures would be a problem.

- 064: OAs built from postcodes that are of approximately equal size would be best.
- 073: 'Mapping' into the postcode geography is essential, but why not grid-based estimates of use in much work using environmental data (soils, geology, pollution etc.).
- 074: Population figures by health authority areas.
- 077: Use of output areas preferable especially if on smaller scale than present EDs (as in Scotland).
- 081: Users should be allowed to specify boundaries of local areas which they would like the smallest output areas to respect (in addition to electoral wards). These might be areas of similar types of housing, or of local projects funded by government or EU.
- 084: Difficult to say there is a good case for sticking with the 1991 EDs for comparison. However, I suspect the choice will be OAs built upon postcodes and I can live with this. I suspect the non-academics i.e. business is desperate for OAs from postcodes. I hope you charge them plenty for doing their job for them.
- 085: Difficult question that relates to issue of comparability of definitions 1991-2001. Use of 1991 EDs/OAs removes one source of non-comparability. But if 'OAs built from postcodes' could be an option for 2001 and all future censuses in frozen form, then the initial 2001 disadvantage would be link it.
- 086: Must have outputs available for civil parishes and wards.
- 111: Whatever output areas are selected, it is vital to have boundaries and explicit links to equivalent 1991/1981 areas.
- 116: Not myself up-to-date with the technicalities or +/- of alternative reporting units.
- 119: Although historical comparability is desirable, it can be maintained at the work level and other methods used for small scale comparisons (e.g. microsimulation & surveys). Output areas should be redesigned to be uniform and the necessary data (from OS) and tools (Martin, 1997) already exist. Spending huge amounts of time and money to digitise new EDs is completely out of question in such a data-rich world, and I wouldn't like to see the 1991 EDline exercise being repeated.
- 132: I think that the basic approach to output geographies set out above is a sound and sensible approach. I have some doubts, however, on whether the idea of defining output areas to avoid all problems of confidentiality will, in practice, be achievable.
- 133: It is very difficult to rank OAs built from postcodes (I view this as a sensible/logical) and 1991 EDs/OAs (I'll always be interested in opportunities to look at change through time), as both are useful and complementary.
- 135: Once again we will be left with an inadequate and inappropriate geographic base. EDs and to a large extent have been proven to be cartographic and statistical junk. When will ONS/OS provide sensible units providing flexibility of output.
- 136: Manchester's 1991 Census geography was such that EDs were comprised of postcodes and this has proved a valuable although time consuming (at the end of the 1980s) exercise. We can now match any service data which has a postcode attached to Census data to build common units.
- 139: Longitudinal work (including identification of 10 year migrants) need comparable boundaries across 4 censuses ideally. Maximum flexibility, please.

3.2 Postcode based OAs

Questionnaire instructions:

Please describe what uses you are likely to make of output areas built from postcodes.

- 001: Use as denominator data in epidemiological research. This is important.
- 002: (1) Data linkage to non-census postcoded data (especially health). (2) Model construction using more uniform areas with more useful georeferences. (3) Mapping and spatial analysis with better knowledge of population locations.
- 004: Much easier to link to other datasets if a postcode based system used. Most of our research with Census 1991 was mapped using DBD; very keen to see any improvement in this area.
- 005: Distribution analysis against postcodes, datasets.
- 007: None - don't buy it.
- 016: Having such output areas would do much to minimise the difficulties of matching postcoded records and census data. This, for instance, is central to much research in environmental epidemiology and health service planning.
- 019: Depends whether the output areas are pre-defined and fixed by OAs or whether the user can at any time create their own OAs from postcode data. The latter would be very useful for example when investigating areas around point sources e.g. power stations, incinerators etc.
- 020: Comparison with other data sources.
- 021: Mapping - boundaries based on unit postcodes UK Borders perfect, as in Scotland.
- 024: Referencing recreation users or travel set points by postcode to an ED or output area.
- 026: Mapping census variables + my own postcode based data.
- 027: Research: in migration and population change analysis and in ethnic analysis, for comparison with 1991 and 1981. Teaching: Classes on use of Census data. Field classes - comparing Census data and field data.
- 028: This would be a huge improvement. Excellent for analysing change in the areas of area-based-policies.
- 029: Linking to postcoded data from other sources.
- 033: Use of existing postcoded data on health/morbidity, utilising health care - direct lineage between individual and aggregated data.
- 040: Mapping.
- 041: Epidemiology - relating census data and other data (such as water quality) to chronic diseases such as childhood leukaemia. This proposal decreases.
- 042: Useful to combine with other MK+researcher data based on postcodes.
- 047: The GRO(S) model is tried and tested and should be adopted across the whole UK without any further question.
- 048: Might be used to construct deprivation index scores for the home address of survey subjects.

- 049: Unlikely to need good specific geographical data. More concerned with getting a finer geography for the SAR's! i.e. - Counties not standard regions.
- 051: Addresspoint would be useful to generate accurate bespoke tables for new areas, but this depends on other arguments on e.g. differencing being won. Postcodes often seem to be held up as in an ideal geography, and this notion should be refused - postal geography is (for good reasons) fuzzily defined, and liable to change. Furthermore, businesses who want to match up customer databases to Census data should be made more aware that the postcodes supplied by their customers probably had a high initial error rate, and are not stable over time.
- 057: Area-based regeneration programmes generally operate on non-standard boundaries. Baseline data construction (for subsequent monitoring and evaluation) and demographic and labour market profiling (for problem identification and analysis) all require Census data on an exact basis. Previous experience with 'best fit' EDs and 1991 data as shown that this would be a preferable option.
- 058: Community analysis. Social analysis.
- 060: Relate to areas of deprivation and healthcare provision.
- 064: Linking in non-census data; mapping; spatial analysis; GIS.
- 071: Mapping of areas by educational cohere, needed to economic prosperity.
- 073: Almost all our work using other data (especially health). 'Translation' between 'geographies'.
- 074: Cancer statistics: - incidence, prevalence survival by various mappings. The postcode is our start point.
- 075: Linkage of primary data to census data.
- 077: Linking medical/mortality data to census data.
- 081: Allocation of administrative datasets (of individual records with postcode attached) to census areas.
- 085: Relate to house prices coded to full postcode.
- 086: Provision of information for other researchers. Poverty/deprivation analysis.
- 090: Service planning.
- 093: Details of uses not planned at this stage.
- 094: Enables comparisons with e.g. insurance based data which uses postcodes.
- 100: Sampling for other survey. Producing deprivation indices.
- 101: Spatial modelling.
- 102: Analysis of spatial difference on Census variables for small areas. Overlay with point data on companies, schools, etc.
- 105: Comparison of residential geographies for ethnic groups.
- 108: Epidemiology: Calculation of disease research, modelling disease incidence and other environmental factors
- 109 Map display

- 111: More precise identification of settled areas within administrative areas. Linking Census data to administrative and marketing data.
- 112: Association with mortality/disease/ stats
- 114: Basis for assessments of health service activity and provision, and health variations/inequality
- 116: The facility is essential - the best method I don't know, but have generally found that the value of documenting change at less than post code sector level is marginal.
- 119: One idea is to create an intermediate geography between OAs and Words. Users who want the finest level of detail (OAs) will have to pay additional royalties. Multi-OAs (intermediate geography) users will only have to pay half of the additional costs involved, while work users can contribute accordingly.
- 124: To aggregate into chronologically meaningful areas (user defined).
- 132: A more accurate approach to assessing the distances involved in work-travel; the capacity to link Census to other post-code based information sources.
- 133: Producing 'bespoke' geographical areas and linking data with postcode referenced datasets (e.g. NHS data) using GIS.
- 135: This will be moderately useful for data integration with postcode based data.
- 139: Modifiable areas, such as ward clusters, or travel-to-work areas for different types of worker, will be very useful in research with a spatial dimension.

3.3 Temporal comparability

Questionnaire instructions:

Please feel free to make other suggestions for solutions to the problem [of temporal comparability].

- 001: Very important to have comparability.
- 002: Virtually impossible to make adequate 2001 OAs which aggregated in all areas to 1991 EDs/OAs. Look up tables from postcode-based 2001 OAs to 1991 best
- 007: Make sure 2001 areas fit within 1991 areas - grid cells only make sense where people live in grids!
- 019: For me the most important thing is to be able to relate 2001 areas back not only to 1991 but 1981 and 1971, using the smallest possible areal units.
- 027: We must have solutions to the time comparability problem in place before 2001 OA statistics are published. i.e. Once the definitions of OAs are known and an 1991 ED/2001 OA look up table is available, the 1991 SAS & LBS can be processed on to a 2001 geography. Then comparisons can be made straightaway when 2001 Census area statistics are published.
- 047: Moving away from EDs - which slavishly seek to equalise enumerators' workloads - allows OAs to be of more varying size: This should allow them to be drawn to match, postcode boundaries (and from them, those which respect topographic 'breaks' such as railways, rivers, watersheds etc.) which should provide greater continuity except in Scotland of course to maximise contiguity with the past Census' small area geography which was (a) totally volatile and (b) meaningless.
- 051: It is possible that disclosure work will show that OAs can be smaller than previously thought, which would make the possibility of 2001 OAs which aggregated to 1991 EDs/OAs (or aggregations thereof) more likely. However this cannot be counted on. Common zones are not likely to be viable over time - they will be disjoint in 2001 and even more so in 2011 etc. Similarly, if we use 1991 EDs/OAs as 2001 OAs, then there will be a good argument to use them again in 2011, 2021 etc. etc. They will become increasingly irrelevant. It would seem better to create new OAs for each Census, and provide good methods to reaggregate/estimate areas from past Censuses in the new geography.
- 057: Using Address Point + GIS/point in polygon, it would be possible to generate 2001 Census data on 1991 EDs/OAs. Access to this data set would have to be restricted - for example, users would be barred from using output for 2001 EDs/OAs - to overcome the potential confidentiality problem.
- 064: Important to recompute past censuses onto a common base for comparison purposes. Choice of base depends on feasibility consideration.
- 073: Why not grid cells for both censuses? - this would nearly avoid the problems and give consistent units forever which 'fossilised' geography will give the least error in the estimate and change?
- 081: Look up tables could include the number of residents and households and residents in households in each intersection of 2001 OAs and other zones.
- 085: I presume these are alternatives. Some approach is important.
- 086: Total population counts and some age breakdowns will be needed for ED level. Won't necessarily need other info.
- 098: I'm not sufficiently versed in the technical details, but comparability (at small area level) is highly desirable.

- 111: Very important to introduce a comprehensive and dynamic look-up from Census, so we know how 2001 Census Areas link into (e.g.) Wards redefined in 2006. Alternatively a fixed geography for the Census and other data sources would be desirable.
- 119: Socio-economic units have a life on their own, they are born, they live and they die out. It is highly desirable to being able to compare them through time, or space, but maybe we should use some more imagination (and research) and try some more radical methods (e.g. pop surfaces or microsimulation)
- 133: A combination of some of the above should meet most needs!

3.4 Protection measures

Questionnaire instructions:

Please feel free to make other suggestions for solutions to the problem [of data protection].

001: Aggregation where numbers are too small.

002: All informed users understand the (severe) limitations of the data. Record swapping preserves consistency of all subsequent aggregations and only has to be done once (pre-tabulation). It results in tabulated figures which are internally consistent and easier to understand.

005: Personally, I find it hard to see why this is a problem.

007: It is not a problem - don't suppress the data.

019: Is this really a problem with basic population data? Perhaps with sensitive cells like annual income, but it would be nice if raw person counts were unadulterated.

027: I hope that research work undertaken either in the academic community or in the Census Office will establish that, for aggregate statistics (tables of counts), threshold populations and broad coding and low dimensionality of tables will be sufficient protection for the data. I suspect, however, that the Census Office will feel that something else will be needed to prove they are doing something positive to preserve confidentiality. Swapping a number of records between areas which could be included in the imputation processes of the one number census project would be the best solution. The number of swaps or % of swaps would be secret and therefore this might be very small indeed.

033: Need to get real - drop all of this and get people to sign/pledge - will not identify or claim to do so.

034: If the cell count is high enough it should be published unadulterated; if it is too low it should be suppressed.

038: Suppression of cells with low counts is common practice in various other European countries: if adopted here this would aid comparability in geographical studies of small areas.

051: It is difficult to make judgements about which of the proposed methods are better than others in the absence of theoretical work examining both the degree of risk associated with the results, and the degree of corruption done to the data. The way in which cell blurring works is easy to explain, and when the table contains large counts (referring to more than, perhaps, 200 people) the error is negligible. However in tables which tabulate data for a smaller number of people, e.g. those which are only just above the 1991 threshold, then the error level is much more significant. This is an important point, given that the thresholds were chosen arbitrarily. If risk assessment algorithms indicate that thresholds could safely be reduced, resulting in a greater number of small OAs, then this problem of error in the data would become of great importance. In the case of cell blurring, cell values will be slightly wrong in some cases, but assuming tables are well above threshold the net modification should be close to zero. It is presumed that roughly the same number of cells will be incremented as will have a deduction. However, the blurring process must prevent cells from becoming negative, and therefore there may be fewer deductions than increments in tables with large numbers of cells with a value of 0 or 1 prior modification. This depends on the nature of the algorithm used. Record swapping similarly leads to some cells having incorrect values, but has several big advantages: there is no need to worry about the 'floor effect' described above; total (of persons, and any other variables that the swapping algorithm holds to be inviolable) will remain exact. If all tables are generated from the same master (containing n% swapped records), then they will be correctly additive: if tables for all the OAs in a ward are summed, they will give the same values as would be found in an independently generated ward level table. This would remove much confusion, as the previous method of blurring cannot ensure this. The level of error in tables generated from a swapped source is not known. In the case of blurred tables, then it can easily be shown that tables with fewer people have larger levels of error. Is it possible that tables generated from a swapped source would have a fixed level of error regardless of the magnitude of numbers? How would this compare to the average error

for all tables in a blurred system? In conclusion: record swapping requires urgent investigation! Work with the SMS has shown that cell suppression causes significant harm to the data, and that was with a suppression mechanism that was too weak! Any 'secure' suppression algorithm would probably cause a lot of harm to the data. Rounding of data would remove a lot of the subtlety that is of prime interest.

- 064: The problems need to be avoided by designing OAs that present no confidentiality risks. Any form data damage reduces the value of the census as a gold standard.
- 081: No measures needed since those imputed or missing add sufficient uncertainty to the census as a representation of the true situation in any area, especially when the fact that response are not 100% reliable or accurate is also taken into account. I think this case would be accepted by Parliament if necessary.
- 084: How can you possibly justify further measures of you are using thresholds! Either the threshold protects confidentiality or it doesn't. It is absurd and stupid to tinker with the data like this. Raise the threshold if you must, but let us have some accuracy.
- 085: 'Small changes to cell counts after table generated' is best if any measure at all is needed. Inaccuracy in data may be improved as much as reduced, given the errors revealed by CVS. Hopefully, the results net out in the totals for rows and columns. Please no suppression or swapping.
- 088: The various sources of 'natural' error should offer sufficient protection. Estimates for this are 8-10 %. It should not be necessary to add further error. Barnardisation should be ditched - especially since it wasn't done properly in '91!
- 098: I feel that changing results is completely unacceptable, and would render research looking for relationships between parameters prone to more errors than were necessary.
- 109: Some low population areas in Scotland are a real problem. Is it possible to ask the populations here for disclosure of some (in fact all) information e.g. on items like language, religion, car ownership, workplace - which are visible in any case to their neighbours.
- 119: Enough is enough with perturbations. What's the point of holding such an expensive exercise and then willingly alter the observations? Why not introduce legal measures to those who claim or even think about identifying individuals? Why not give residents the right to object to users when they felt their data is misused? Why is there the underlying assumption that the users will try to identify individuals, in the first instance?
124. The absence or relational absence of a person with a particular contribution is in itself important information. This should be avoided and, if necessary random change to a small proportion of characteristics made.

3.5 SARs

Questionnaire instructions:

Please feel free to make other suggestions for SARs in 2001.

- 007: A 100% SAR? Life would be a life easier - confidentiality isnot a problem.
- 009: Some SARs questions contained too much detail - e.g.: occupational classification whereas questions or more specific importance (EDUCLEVEL) were poorly represented in the SARs - more quality in some of the questions. Other questions - a balance needs to be achieved.
- 019: I have not used these in the past but I may use in the future.
- 020: I find the pre-tabulation of data into classes confusing e.g. the uneven class sizes employed for distance to work, and the uncertainty of where intermediate value are placed e.g. does 1.5 km belong with 0-1 km or 2-3 km.
- 025: For many non-geographers like myself variable detail is more crucial than geographical detail.
- 027: My advice would be to stick with a Household and Individual SAR broadly comparable with 1991, so that case-building effort could be focused on the Local SAR proposal which will need careful confidentiality research. This would simplify the task of the Census Microdata Unit.
- 028: SARs 91 was extremely valuable and the 2001 SARs should go further in terms of geographical disaggregation.
- 034: SAR geography needs to reflect the policy process which occurs at unitary level if sub-district geographies are constructed they should as far as possible follow internal LA planning areas (all LAs have these).
- 064: Important to continue the SAR tradition. The variables should be comparable over time and the SAR sample applied to other censuses.
- 065: In addition to the existing 2% Individual SAR, there should be another one with only region coded but with the most detailed labour market data fully coded (as in the HH-SAR).
- 071: Again I think a software solution may be possible.
- 081: Individual SAR with LA geography with much less detail on some variables (occupation, industry, country of birth...) and more derived household variables. For those interested in categories with few people in them, the 2% SAR will never be good enough for local areas, and the 1% SAR can be used for national relationships with other variables.
- 083: These should be larger individual SAR, e.g. 5 %.
- 085: 'Sub-LA geography' not as useful as adding area-type coding.
- 086: The migration variables need to be modified. Distance of move is meaningless for a very large area. We need to know whether someone has moved from other parts of the County, District, or Region or outside Region.
- 097: We need bigger SARs especially for ethnic minority research.
- 113: A household-level non-hierarchical SAR with more geographical detail than the 1991 hierarchical SAR
124. The more local geography the better! Substantial gains are made in extent of data compatible with confidentiality constraints (particularly in light of answers in 3.1.6).

3.6 Migration statistics

Questionnaire instructions:

Please provide comments and suggestions for the production of Migration Statistics from the 2001 Census.

- 019: I have not used these in the past but I may use in the future.
- 027: (1) Produce a comprehensive set of migration area statistics at national, regional, local and small area scales. (2) Produce a comprehensive set of migration flow statistics as origin-destination based arrays in easier to use formats. (3) See the review paper presented at the Third Workshop Planning for the 2001 Census for many more suggestions.
- 047: The rounding or suppression of small flows was totally unacceptable, and some other form of confidentiality guarantee should be applied (if necessary).
- 051: The tables produced for the SAS/LBS in 1991 were of very little use, and should not be replicated in 2001. The one number census should allow for the production of much more useful tables.
- 064: A finer level of output geography should be considered. 100% coding.
- 085: All these mentioned above, - timely delivery of SMS; - easier to use RMS; - outmigration data in LBS. But in migration 'not stated' must be tabulated separately and not subsumed within ward moves; - no suppression in SMS; - more variables in SMS (income, occupational group).
- 086: Need to ensure that student usual addresses are correctly coded. As an enumerator in 1991 it was aware of cases where migration from term time addresses in 1990 to student's home address in 1991 would have been recorded.
- 088: The 1981 & 1991 SMS were unusable due to the effects of suppression. This can't be allowed to be repeated for 2001. More migration information with LBS/SAS desirable. Don't make the assumption that persons with origin not stated moved within the ward of residence
- 102: Agree entirely with need to report out-flows as well as in-flows.
- 111: Need for a simpler (matrix) format for the data, where it is clear what the flows between areas are. Need to reduce the amount of data suppressed. 1991 SMS for all its faults was easier to use than 1981 SMS.
120. Low rates in population subgroups limit the interest of one-year migration. I would like to see emphasis on integrate of 2001 Census into the LS to examine migration issues..
124. Between, at highest level, postcode sectors is desirable. (Local migration is likely to be an important determinant of mortality, particularly in deprived areas in Glasgow.
130. At present limited use because no birthplace/ethnic variables. I would use Irish identified migrants (plus other characteristics - age, gender, employment etc.,)

3.7 Workplace statistics

Questionnaire instructions:

Please provide comments and suggestions for the production of Workplace Statistics from the 2001 Census.

- 019: I have not used these in the past but I may use in the future.
- 020: Could they be cross-tabulated by Ethnic Group.
- 021: Have used SWSTAB to extract information. Would like software that is a bit easier.
- 027: Two issues need addressing for 2001: (1) Given the intention of the Census Office to code all questions at a 100% level, what confidentiality measures should be applied. I would argue that Sampling is still the best option. (2) The accuracy of look up tables based on the workplace postcode needs to be addressed. There were deficiencies revealed in these LUTs associated with the 1991 Census. (3) Careful attention is needed to code the workplace location of persons whose main workplace is their residence.
- 028: Clearly the coding of postcodes into EDs words could be more accurate. This is essential for the analysis of spatial policy. This process would be properly documented and available to researchers.
- 044: Does this mean workplace, i.e. question I21-I28 or travel I29?
- 047: There are grounds for introducing 'additional protection measures' if the 10% sample process continues, because there is no evidence of public concern over this form of data; if 100% processing was introduced then some form of protection would be unacceptable 'price to pay' so long as it was not the rounding or suppression of cell counts (with small values).
- 051: There were many problems with these statistics in 1991, due - as far as I understand it - to difficulties in allocating postcodes to workplaces and mapping from those to EDs. It is to be hoped that the one number census procedure recognises these problems and has plans about how to overcome them.
- 057: Use of Address Point would considerably enhance the accuracy of the 'distance-to-work' variable. It would also be helpful if the distance ranges used in the output were increased by (say) 2, especially at the lower end of the scale. I would also be interested in more tabulations using the distance ranges - cross reference by mode of travel, occupation (major groups), age and hours worked would be extremely useful.
- 064: 100% coding needed. Finer level of output geography.
- 065: Home-based work is a growing trend and tables/statistics need to address this topic more fully in future, in particular to differentiate between: - works mainly at home (no transport); - works from home as a base (use transport).
- 085: Might have been interesting to know % hours worked at 'main workplace'! E.g. average number of days/weeks that commute is made. My indirect use of the dataset is crucial - in terms of using commutershed areas (e.g. local labour market areas) for population and employment change analysis.
- 088: - More detailed worker/workplace tables as part of LBS/SAS (e.g. home workers) by industry, journey to work by industry etc.). - 100% codings but not if it results in suppression for interaction data.
- 111: Very important to introduce a table of journey-to-work by ethnic group.
- 132: I am sure that a significant issue for the SWS is how households actually fill in their Census forms with accurate postcodes of workplaces. The more that can be done at this stage the easier subsequent processing problems will become.

3.8 Longitudinal study

Questionnaire instructions:

Please provide comments and suggestions for the linkage of the 2001 Census to the Longitudinal Study.

- 007: We can't use it properly while it is locked away.
- 019: I have not used these in the past but I may use in the future.
- 025: I anticipate using LS data in the future and thus regard the addition of records from the 2001 Census as important.
- 027: (1) Very important (unique) data set which must continue to be supported. (2) Still needs considerable investment of time by ONS, LSSP and user which limits number of projects that can be supported. (3) Investment in better software to speed turnaround in table production and vetting should be encouraged. (4) Is there any possibility that selected tables produced for projects could be released for general use?
- 031: It is essential that the 2001 Census is linked to the LS as quickly and as fully as possible.
- 033: Again there is a problem of non-disclosure of small counts - provide safe analytical environment which gives results from modelling (e.g. multiple modelling) but not access to original data.
- 047: The LS provides the only means for answering questions such as 'Are the young people of area X today found it easier to get a job /home of their own than did similar young people 10 or 20 years ago?' - and there are many other (especially epidemiological) enquires for which a large scale cohort needs to be followed through space and time.
- 048: The importance of the LS lies in its size, and consequent statistical power, and its representativeness. In terms of health analysis (mentality and long-standing illness) it is essential that 2001 census data linked into this unique resources.
- 049: Greater comparability with the SARs - i.e. - it would be very helpful to achieve linked household (structure) SARs data to trace formation/education/disabilities of households/household structure.
- 064: The LS is an expensive anachronism of limited value outside of medical research. The SAR is far more useful.
- 065: The LS becomes an increasingly valuable dataset the longer the time span covered and as we gain experience from the SARs. Linked data for 1971-2001 will permit complete work histories for people aged 25-55 years to be produced and analysed.
- 070: The LS is now coming truly of age. It is possible to investigate the relationship between childhood circumstances (e.g.) outcome of pregnancy, marital breakdown, family formation, patterns of employment, social mobility in younger adult life. This needs to be extended to older ages. The inclusion of health question now means that mortality data can be used more quickly. The investment will pay greater deviance the longer the study continues.
- 085: Yes, must have 2001 Census link up - hopefully as few as possible will need to be imputed or estimated! Ten-year transitions very valuable, especially being able to study geographical and social mobility at the same time.
- 098: The LS is a unique resource, and must be kept 'live' - if the 2001 records were not added it would become a historic archive only - and not an evolving dataset.
120. As soon as possible! Better user access to the LS would be welcome.

139: Four censuses offer greater possibilities for a 30 year perspective on individual [lives], for comparison of cohorts over 20 year spans of the lifecourse, for comparison two generations in the same family, and for exploiting the accumulation of evidence on rare events.

3.9 Boundary data

Questionnaire instructions:

Comments on the Digital Boundary Data/Suggestions for 2001 Outputs.

- 004: Mapping census analysis results makes interpretative so much easier and any improvements in this service would be welcomed.
- 007: Use 1991 EDs then you only need sub-ED boundaries.
- 011: Excellent scheme which must be continued and even expanded if possible to include background digital data such as roads, urban areas, rivers etc.
- 016: We have found it easier to download ED boundaries for individual counties from MIDAS and process them locally than use UK BORDERS.
- 019: I have not used these in the past but I may use in the future.
- 023: We need two kinds of software: (1) 'Quick and dirty' so that anyone can draw quick, rough maps in a two minutes. (2) Window-based user-friendly software so that non-specialists can produce plots of minimal standards for publication without having to spend weeks learning the package.
- 024: I would hope that another Quality Assurance exercise would be undertaken. This should be done for ED/OA centroids, as well.
- 027: To supplement the general access software (UK BORDERS), the service needs to develop a library of common coverages in standard formats which can be quickly downloaded by users. This point also applies to the SMS and SWS.
- 028: Essential for defining areas for spatial policies.
- 034: The health authority problem needs to be addressed, county count level boundaries would be helpful.
- 047: The production of paper and microfilm copies is valuable BUT ONLY if they are of innumerable superior quality than in 1991 - which were very poor reproductions technically, and were monochrome copies of maps for which colour separation was indispensable for correct interpretation!
- 051: The boundary data are very important to my work, however there are a number of suggestions that could be made. The boundaries were detailed - this is obviously a good thing, but for any region of interest larger than e.g. a local government district they were *too* detailed. Whilst it was entirely possible for the user to generalise data, it would have been preferable if this had already been done centrally to some extent. (Perhaps this is a criticism of the data distributors rather the data itself). I'd suggest that a number of boundary sets be generated (with obvious caveats that if you try to merge boundaries from different sets then you are likely to get mismatch problems at the borders). E.g. A set of boundaries, down to district level that are envisaged for use when the region of interest is either national or close to national; a second set to ward level suggested for use at roughly standard region level, and a third set, as detailed as the 1991 ED boundaries, intended for localised use. The cost implications of this should be small - the generalised boundary sets can be generated automatically (albeit with systematic (automated) error checking) from the most detailed set. The only costs would therefore be a few man-days, and the disk space to store multiple boundary sets. The licensing arrangements in 1991 were restrictive over reproduction of the boundaries. It is very likely that the World Wide Web will be the medium of choice for delivery of 'papers' containing maps, especially if the maps need to be in colour. The licensing of the 2001 boundaries should consider the role of the web carefully. The availability of an 'official' generalised set of boundaries might assist in this.
- 064: Digital boundaries are an essential census output in the GIS age. They need to relate to 2001 OAs.

- 073: I see no point in using overdigitised boundaries - there is a clear market for generalised DBD which would meet many teaching and research needs.
- 085: Only indirectly interested in terms of possible area aggregations and analyses that can be produced by others, which I may use.
- 088: Boundary datasets must be released earlier. - Quality assurance important + libraries of generalised boundaries.
- 093: Absolutely useful to link the Census statistics output with digital boundary data.
- 102: Libraries of pre-selected boundaries would have been very time-saving! Importance of generalisation - since some boundary files are very big! Ensure availability - different formats.
- 110: Try to get rid of more errors in the boundary files (i.e. problems caused by rivers etc).
- 111: Vital to have access to up-to-date digital boundary data. Also very important to have inter-censal boundaries (from OS boundary line (?)). UK BORDERS had been useful, but would be more useful if pre-created geographies could be accessed (e.g. by FTP) rather than building up areas from scratch in each access.
- 119: Why should digital BD held both in MIDAS and in Edinburgh? Even if they have to be duplicated why should they be incompatible? Central access or automatic update of multiple versions is trivial in an Internet era, yet the people in charge have never heard of compatibility and comparability within thesame dataset?
- 135: 1991 was clearly an improvement on 1981 (but then I would say that). WWW based interfaces to these data are clearly the future.
- 137: Please add Northern Ireland boundaries and do not let many more changes take place in Scotland in particular. However, Northern Ireland boundaries in terms of government areas must allow for the recent changes in boundaries since the 1991 Census.
- 139: Boundary data are used in combining areas in the LS to look for meaningful area patterns in poor health.

3.10 Look up tables

Questionnaire instructions:

Comments on the Look Up Tables/Suggestions for 2001 Outputs.

- 007: Use 1991 boundaries - we need to have a look up table to 'county wards'.
- 016: Please ensure there are appropriate quality control procedures. The ED/PC directory still has some gross errors in it.
- 019: Ability to compare 1971, 1981, 1991, 2001 census areas to as small an area as possible.
- 027: A well organised family of look up tables (or general relational database) will be essential in 2001. It should include: (1) OA/PC directory updated as PCs change; (2) OA/wards, LGDs updated as wards/LGDs change; (3) 1991 ED/OA to 2001 OA look up table so that change over the 1991-2001 can be studied.
- 047: A connected 'best fitting' effort to create Tables from 2001 OAs to 1991 EDs/Wards is essential - and a vital follow through of this will be to then generate Tables of 2001 OAs to all the other pre 2001 geographies.
- 051: These are very important. The PC/ED file was very useful, although difficult to use: comma separated fields would have been nice! However, their requirement in future will obviously be dependent on the base geography chosen.
- 064: Look up tables are a fundamental census output.
- 077: ED/Postcode file for N.Ireland is not checked or validated - need to bring N. Ireland data availability into line with GB.
- 085: I will need ability to aggregate SAS to higher-level non-standard geographies, e.g. Inner/Outer City, Urban/ Level areas, FR zones.
- 088: The look up tables must be kept up to data and also be subject to QA.
- 111: Look-up tables are vital. There should be more of them, linking Census areas to pre and post - 2001 administrative/postal areas.

3.11 Flexible output system

Questionnaire instructions:

Comments: Please feel free to make suggestions.

- 006: This would be the single bigger improvement, from my point of view.
- 019: Very important to study point sources or interest, to be able to define any area anywhere.
- 027: Users agree on the desirability of this as a goal. The technical means of achieving the goal are under investigation. The Questions that need answering are: (1) Who pays for the tables/outputs? (2) How much can the user do and how much do the Census Offices do? (3) How is the queuing problem to be solved? (4) How are disclosure risks to be assessed? (5) How is the history problem in table release to be solved?
- 033: I rarely want the individual returns for themselves but as input to analyse all therefore interpret parameters - 'safe data storage seems a better' way forward for this sort of use.
- 034: Too expensive and much too slow. Cost is crucial.
- 047: If arrangements can be devised to greatly reduce the 'entry cost' of Special Tables. Then 1991 data should be supplied through the same service/facility.
- 051: I am in a slightly unusual position in that the difficulties surrounding this are of interest to me more than the actual end product.
- 057: This would be particularly important for work on non-standard areas used in urban regeneration, economic development and regional policy (see above).
- 064: The SAR data provides a zero cost alternative. Users should be allowed to design their own census output geography.
- 065: I used Special Tabulations from the 1981 Census, before the SARs release, as part of my work in a central government department. This option becomes more valuable now that the SARs can be used to test + check table design and analysis problems before formally ordering a Special Tabulation.
- 073: If machinery was available for customised analysis, the standard area tables could be less exhaustive.
- 081: I would make use of flexible census output. ONS currently prices the request for flexible output so that costs cannot easily be included in research proposals. The main problems are high cost and long delay after ordering, both of which would be reduced with an appropriate software system.
- 083: A maximum amount of tabulation should be available on CD-ROM - much more than is currently produced.
- 088: If the disclosure risk issue can be addressed/solved - could an academic site take on the role of 'safe setting' for this service?
- 111: Such a system would be very useful, enabling me to generate the kind of tables I have produced using the SARs, and avoiding the problems of small numbers from the 1% or 2% samples. The availability of such a facility could greatly simplify Census output, reducing the cost, by enabling people with more specialist needs to produce their tables without adding to the vast number of pre-specified tables in the printed volumes and LBS/SAS.
120. Generally SARs greatly reduces the need for special tabulations.

133: Not used this - but it might be sensible to liaise with NHS and produce a table with long-term illness and/or general health, with a sensible number of covariates (age categories, sex, and a few socio-economic variables).

PART 4:

THE ESRC/JISC CENSUS PROGRAMME FOR THE 2001 CENSUS

4.1 1991 registration procedures

Questionnaire instructions:

Comments on registration procedures.

- 002: Difficulties encountered locally due to use of computing staff with no census knowledge or interest as site representatives!
- 011: Nightmare bureaucratic process which must be made simpler. (I know steps are being taken).
- 015: Institution registration took months for DBD.
- 019: Would be easier if you only had to register once to access all data, instead of separately for each dataset and each institution.
- 021: Lengthy.
- 022: The different forms for GB & NI cause complications.
- 024: Rather awkward in a teaching context (in all MSc and undergraduates have to fill out the forms, too).
- 027: (1) Too many separate agreements which need to be consolidated into one or two comprehensive agreements for users and institutes. (2) Maximise the availability and use of electronic versions of registration forms.
- 029: A single registration for all census procedures is highly desirable.
- 034: I was originally asked to get the vice-chancellor to personally fill out and sign the universities registration form. This is way over the top. Despite the existence of detailed terms and conditions in the registration documents, it was completely unclear how much census data you could publish in an academic publication without first submitting it to OPCS for approval.
- 040: Too complex - we should move to an one form system.
- 043: Problems stem from (1) a dependence on local university representatives who may have little interest in census analysis, and (2) the plethora of registration forms required (one for each product).
- 045: Complex because my institution not aware of its responsibilities.
- 047: Now quite good - but previously a bit off putting.
- 051: Satisfactory, I suppose, in that I did manage to register. Unsatisfactory, however, in the sense that this required signing multifarious different forms, often providing the same information (name, address etc. etc.).
- 057: Generally OK, but a bit cumbersome having to go through the University representative.
- 058: Laborious for undergraduate and graduate classes.
- 064: Far too cumbersome, complex.
- 066: Rather bureaucratic, poor for non-UK users.

- 068: Had to register for each data - set separately and go through MIDAS site rep, both of which took time. New proposals sound encouraging.
- 073: On-line and WWW service good, but arrangements for registering large number of students must be made simpler if these data are to find full teaching use.
- 077: Seem to remember having to fill out multiple forms for SAS and SARs, also for GB and N. Ireland.
- 084: Manchester is fine and helpful. Very good. Edinburgh is terrible, too many forms and ridiculous amount of witnesses, signatures etc.
- 088: Too many forms. Procedures problematic for casual user (e.g. undergraduate wanting word profile).
- 102: Time consuming!
- 105: Cumbersome process. Needs stream lining.
- 119: Maybe somehow exaggerated with students sometimes. Who would want to use census data without the MIDAS support? Also, somehow complicated procedures are involved when registering for different datasets. Is this necessary?
- 135: The registration procedures have been proven to put people off using the data set. Potential users are unclear as to what registration forms to fill in what their rights are etc.
- 139: The University rep is virtually useless, never available, and not regarded by her manager as having any responsibility for research.

4.2 2001 registration procedures

Questionnaire instructions:

Comments on proposed registration procedures.

- 002: One-stop registration for as many of the datasets as possible has to be a good move!
- 007: It should be illegal to restrict access to census data.
- 011: Finds a lot more sensible and easy to use
- 019: One stop registration better.
- 021: Good to make more use of the web.
- 022: Single agreement for institutions to sign is an important, as is a single agreement for users to sign.
- 024: Better than current system.
- 025: The 'one-stop' system sounds excellent.
- 027: I don't see any alternative to such systems unless the Census Office decide to release the data under a different costing/ownership model.
- 034: What about boundary data, will this be included in the single registration.
- 047: Sounds like a big improvement!
- 051: The necessary technology to allow simple semi-automated on-line registration is already in place or nearly in place, and everybody seems to be agreed that web based registration is desirable.
- 068: I think these are very good proposals if they are implemented properly.
- 073: I worry about the feasibility of using local SAS in teaching large classes. Some arrangements for benchmark, 'public' data needs to be in place to assist software/methodology research.
- 084: If the above is done it would be very helpful. But when we are going to leave a class form for the boundary data.
- 085: One-step system best. Realise that LS must be different.
- 088: One form for all 1971, 1981, 1991 & 2001 'safe' datasets + on-line registration. Also - registration to cover boundary data - make it possible to use boundary data as part of interface to the data.
- 091: Welcome development of on-line access to data in a safe environment e.g. micro data with small geographical areas and contextual detail.
- 119: Great ideas, especially the one-agreement proposal and the on-line and web-site access!!
- 135: Should be a great improvement. I would also like to see on-line registration. I am aware of the legal properties of the signature but steps should be taken to allow some form of secure on-line registration. Possibly involving an encrypted signature sent [by email].

4.3 Help from the Census Programme

Questionnaire instructions:

Comments and suggestions on the help/advice services provided by the Census Units.

- 001: More staff required for support in this service.
- 002: Midas census gopher was good, but can be hard to find associated useful information. A well-written web-based interface is essential.
- 011: Particularly high praise for UK BORDERS (Alistair Towers before he left, then Anne Davidson!)
- 016: I have used both CDU (MIDAS) and UK BORDERS. I would rate the CDU staff as excellent and the UK BORDERS service as OK but nothing special.
- 019: Help provided by individuals has been very good, but on-line help such as gopher is inadequate as you have to know what is there and where before you can even hope to find it.
- 020: Workshops and seminars need to be more available outside academic term-time.
- 021: Training courses needed in Scotland.
- 024: Ratings: gopher=1; www=2; phone/email 1-2; dataset manuals 1-2; software manuals 1-2.
- 027: Excellent service/help provided to me at all times.
- 034: SASPAC needs to be made much easier to use and should included an integrated mapping model. If this occurred then census data would be much more widely used. The 1991 census data was not as widely used as it should have been.
- 045: Assumptions made about my level of computer literacy which have been unrealistic to the point of absurdity.
- 047: There can be no doubt that the Co-ordinator has done a great deal in helping and encouraging users. Outstanding job and MIDAS haven't been for behind, with one or two of the other participants less successfully helpful.
- 057: User Guides etc. were excellent, but the manual provided to show how to extract data on-line was poor - badly thought out (from a user's point of view), lacking certain key details, and with low production values. Hopefully it will be better next time!
- 071: I think a more integrated approach to all the issues around the world would be better. You could use people outside of the Census units more perhaps, to give talks/demonstrations, get a register of volunteers (check them!).
- 084: Manchester is generally very good to outstanding. Edinburgh has been poor at times, especially with its development of the interface.
- 088: Better integration of documentation & training.
- 102: Manchester - excellent. Edinburgh - help very good despite software + data problems.
- 108: Better searchable web pages. I understand they are under construction.
- 119: Now, this is tricky. The Leeds University authorised personnel's help is outstanding. However, it is clear that the help desk of MIDAS is a bit slow in responding, not necessarily implying that the people don't work hard, but I suppose the staff is not enough. As for the UKBORDERS, no comment!

135: I feel that for 91 the approach was satisfactory but would it be in 2001? The key to increasing [the] number of users and effective use of the data resource lies in amalgamating access to data with education about the data set. The data distributors must make [sample] data [available on the Web].

4.4 Documentation and training

Questionnaire instructions:

Comments and suggestions on documentation and training for the 2001 Census.

- 011: No. 3 Census Programme Packs - some were fine, some really difficult - especially the multiple publication and manual produced by the Data Archive!
- 020: On-line information needs to be more user-friendly i.e. interactive, and not grouped by data set.
- 022: Note my role is support - not used documents for my research. documents and training are important/highly desirable for research users.
- 027: All training materials produced by project/researchers must fit into a post-production dissemination strategy. Provision via the Web and maintenance by an ESRC/JISC Census Unit is essential and should be a requirement in 2001.
- 051: The User Guides were generally useful, but could be more so. All User Guides or equivalent should be freely available on the web. Some of the technical information was incorrect (at a fairly pedantic level, but irritating all the same). For raw data their should perhaps be Programme Guides to supplement the User Guides, which would contain all relevant information about various codes used, the edit processes done by ONS - especially in 'difficult' cases - and so on.
- 064: It should be www based, automated, on-line.
- 065: The Census Offices' documentation needs to be much more detailed than in the past (1991) now that academics have access to microdata.
- 068: More regular one day courses for the more popular workshops, such as the introduction to the Census. Should occur more often than once a semester.
- 077: Census Office User Guides are not well written for those with no familiarity with the system (i.e. SASPAC). Could some training materials (i.e. as used in Workshops or large courses) be provided either in published form or on the Internet, for people who cannot attend for financial (travel) reasons?
- 084: Please restore the gopher at Manchester and maintain it. (Likely I know).
- 088: Better training/documentation for librarians, the people who provide access to the CD-ROM products.
- 098: Would expect 'on-line' source to be come much more important.
- 102: Good manuals for using SMS, SWS, LS, SARs etc. as well as SAS/LBS very important.
- 109: The MIDAS documentation was first class.
- 119: Just keep the standards up!
- 133: MetaC91 was outstanding and has provided a role-model! BIRON not so important in terms of using Census data. SAR codebooks were very good, but an index and page numbers would have been useful. I suppose these could be put up on the WWW for 2001?

4.5 Methods of access

Questionnaire instructions:

Comments and suggestions on methods of access for the 2001 Census.

- 002: Web access for authorised selection and download essential.
- 011: Logins via ftp & web very useful and the best way to download data!
- 022: Again, my role is support - so I have not used the 1991 for my research. On-line access to server is essential, and preferable to transfer of new data to local computer.
- 025: CD-ROM's are becoming an increasingly attractive mechanism for accessing data
- 026: Web access please.
- 027: (1) Virtually all of my use has been on-line to the MIDAS service for extraction and for analysis. I have used both MIDAS and local computer facilities depending on the task. I regard MIDAS as my Census archive and transfer (via FTP or Rapid Filer) only the specific data being worked on. (2) The importance of Web access in the future will be that this will be operating system independent (at both server and client end), so that many more users - at present deterred through lack of Unix knowledge - can be encouraged to access the data.
- 045: Still having basic problems because of mis-match of computer-literacy assumptions.
- 047: It should be possible for some tables to be available on line from LS (e.g. if every cell is of a certain minimum size, or has been extensively 'adapted' to preserve confidentiality) - so that users can experience before requesting definitive tables.
- 051: Most of my access has been on-line via MIDAS. A web based interface would be nice, as long as it does not prohibit systematic running of different batch jobs etc. For transfer of data by physical media, CD/DVD should be a priority.
- 057: Internet access would be a great boon, especially if it can make use of GIS technology. I have an (idealised?) image of sitting here in front of my PC, looking at a map of the UK, and being able to zoom into an appropriate scale to define my areas of interest, to select the data sets/tables/variables for these areas, and then to save or transfer the file straight back here. Or would this be too expensive on interface to develop?
- 064: www/ftp access.
- 071: Standardise X use for Midas, and give recommended environments (login, cshrc etc.).
- 073: Must be by FTP from WWW access or CD for large files.
- 083: Maximum availability of data on CD-ROM and via MCC.
- 084: I'm quite happy logging on it -the less the data much more flexible is its use.
- 088: We should be able to use boundary data as part of an interface without having to check that the user is registered for the boundary data. Better integration of meta-data with data extraction system.
- 110: A more accessible system of transferring data easily into a variety of spreadsheet formats that can be used in desk top statistical packages.

132: While the questions above have some relevance to previous behaviour that behaviour was conditioned by the computing environment in which individuals operated. It is important to grasp new opportunities for the 2001 Census Web transfers of data, better CD-ROM [data collections].

135: web web web WEB WEB WEB!!!

139: LS tables can come in various forms: paper, machine readable grouped data, spreadsheets.

4.6 Area statistics software

Questionnaire instructions:

Please indicate which software packages are important for your research using census data?

001: Standard

002: SPSS, ARC/INFO, EXCEL, Own programs

004: SASPAC, SPSS, EXCEL

005: ARC/INFO, IDRISI, EXCEL, FOXPRO

006: SPSS, EXCEL, ESS, Maps

007: C' program

009: SASPAC, SPSS, EXCEL

010: Atlas, MapInfo GIS

011: Spreadsheets, Mapinfo

012: SPSS

013: SPSS, GLIM

015: SASPAC - text files for input into spreadsheets (EXCEL, DBASE)

016: SASPAC, MINITAB, SPSS, DBASE

019: SASPAC for extraction of raw data for processing using my own programs.

020: SASPAC

023: SASPAC, SPSS

024: SAS

026: SASPAC, SPSS

027: SASPAC (Unix), MapInfo (Windows), SPSS (Unix & Windows), EXCEL

028: SPSS-X

029: SASPAC, EXCEL, BD, SPSS, ASCII

030: SAS

033: MINITAB, ASCII

034: SASPAC, SPSS, SYSTAT

038: By file transfer, manipulated in EXCEL, mapped via SUPERMAP.

039: SASPAC, SPSS, Lotus 1-2-3, EXCEL

040: SASPAC, ARC/INFO

041: ACCESS, STATA (statistical software)

043: SPSSX

045: MINITAB

050: SASPAC, SPSS, EXCEL, GLIM

051: SASPAC. I would have used a spreadsheet format, although EXCEL is good at importing text.

057: EXCEL/ACCESS/SPSS

058: EXCEL

062: SPSS, LOTUS, LIMDEP

064: Own

067: SPSSX - SPSS format; Limdep - ASCII format

073: ASCII(TEXT) and ORACLE/INGRES

075: EXCEL, MINITAB, SPSS, GIS

076: SASPAC

077: SPSS, EXCEL, also text files for other software

081: SASPAC 91

082: SASPAC, EXCEL

084: EXCEL, ARC/INFO, ASCII, - comma delimited. You must maintain ASCII as it is versatile

088: Web based + reformatting tools. Data held in database to facilitate variety of interfaces being developed.

092: SASPAC, EXCEL, MINITAB

093: Spreadsheet and statistical software (e.g. SPSS, MINITAB).

094: EXCEL/SPSS

100: SASPAC

102: SPSS, EXCEL, ACCESS

104: EXCEL, MINITAB, MATRIX FILES

105: SASPAC

106: SASPAC - horrid to run, what about ODBC interface?

109: Used Quanvert via MIDAS. Could be more flexible.

110: SASPAC - prefer portable spreadsheets

- 111. SASPAC, own programmes reading ASCII files output by SASPAC
- 112: ACCESS, EXCEL
- 119: SASPAC, EXCEL,
- 120: SASPAC
- 123. SAS (statistical software) for analysis EXCEL - for spreadsheets/presentations
- 132: Any spreadsheet or database format, but there will always be a need for a “back-up” ASCII file that can be accessed through user written software in a recognised programming language.
- 133: MLn, Minitab, SPSS, S-PLUS
- 135: Use standard database/spreadsheet formats. Use XML for metadata.
- 136: SASPAC, SPSS
- 137: WWW
- 138: On CD-ROM
- 140: SPSS

4.7 SARs software

Questionnaire instructions:

Please indicate which software packages are important for your research using census data?

001: Standard

006: SPSS, EXCEL, ESS, Maps

007: C' program

009: SPSS, EXCEL, Data sheet I SPSS, Spreadsheet for mapping

010: Atlas, Map Info GIS

012: SPSS

020: SPSS

025: SPSS/System files

027: USAR, SPSS (Unix & Windows)

028: SPSS-X

029: SPSS, EXCEL, ASCII

033: MLn, SPSS, BMDP-ASCII

034: SPSS

039: SPSS

040: USAR

043: SPSSX

049: SPSS on-line access

050: SPSS, GLIM

051: USAR

064: USAR

065: SPSS

066: SAS, SPLUS, MLn

068: Extracted data using SPSS but have had to convert it to be readable in my preferred package (STATA)

071: ASCII, DATA MATRIX

081: SPSS

082: SPSS

083: SPSS

088: SPSS & SAS + any fast tabulation system that might become available to the community. Will ONS allow web based access?

091: SPSS SAS

094: EXCEL/SPSS

105: USAR

110: SPSS - prefer portable spreadsheets

111: SPSS, ASCII output of derived data files.

119: USAR, SPSS

120. SPSS

123. SAS (statistical software) for analysis EXCEL - for spreadsheets/presentations

130. Quanvert

132: At the moment I am happy with the SPSS approach.

133: MLn, Minitab, SPSS, S-PLUS

136: SPSS

4.8 Interaction software

Questionnaire instructions:

Please indicate which software packages are important for your research using census data?

001: Standard

007: C' program

027: SPSS, GLIM, SMSTAB, Various Spatial Interaction programmes

028: SPSS-X

039: SPSS

040: SMSTAB

043: SPSSX

050: Quanvert, excel

051: SMSTAB, SWSTAB

057: EXCEL/ACCESS/SPSS

064: OCON

081: SPSS

085: SMSTAB

088: Data in a database initially until other software options can be evaluated.

094: EXCEL/SPSS

102: Own software, SPSS, EXCEL

110: SPSS

111: Quanvert, SMSTAB/SWSTAB. My own FORTRAN programs using data output from SMSTAB/SWSTAB.

132: Same comments as above - for wider use some developed accessing software is needed but with an ASCII backup to allow some more imaginative use.

133: MLn, Minitab, SPSS, S-PLUS

4.9 Longitudinal Study software

Questionnaire instructions:

Please indicate which software packages are important for your research using census data?

001: Standard

006: SPSS, EXCEL

007: C' program

013: SPSS, GLIM

023: EXCEL, SPSS

025: SPSS/System files

027: SPSS (Unix & Windows)

028: SPSS-X

030: SAS

033: SPSS, MLn ASCII

034: SPSS

039: SPSS

049: SPSS, Printed hard copy cross-tab to specification

065: SPSS

091: SASPAC MLn

094: EXCEL/SPSS

095: SPSS, STATA

111: SPSS, SAS

120. SPSS

133: MLn, Minitab, SPSS, S-PLUS

137: WWW

139: SAS, SPSS, STATA, MLn, FLEXTRACT, LIMDEP, ARCINFO

4.10 Boundary data software

Questionnaire instructions:

Please indicate which software packages are important for your research using census data?

001: Standard

002: ARC/INFO, Mapinfo, IDRISI, Own programs

004: Mapinfo: hence, any related data format.

005: ARC/INFO, IDRISI

007: C' program

009: Exported digital data

010: Atlas, Map Info GIS

011: Loaded into Mapinfo, Arcview, Arc/Info

015: ARC/INFO format

016: ARC/Info GIS

023: Mapviewer

024: EDB, SURPOP

027: MAP INFO

028: ARC, MAPINFO

029: ARC, MAPINFO

034: GIMMS, MAPINFO

038: SUPERMAP

039: Mapviewer, Mapinfo

040: DBD 91

041: Arc/Info/MapView (for my sins!)

051: UK Borders, DBD-91, ARC format

057: MAPINFO

058: GIS - ARC INFO, ARC VIEW

064: ZDES, ARCINFO ARCVIEW - PC/UNIX

073: ARC/INFO

077: ARC/INFO, ArcView

- 084: ARC/INFO Export Format.
- 088: Web based + reformatting tools. Also - internet based map serves software for delivering census maps are the web.
- 093: ARC/INFO; MAPINFO; etc.
- 094: EXCEL/SPSS
- 102: Map/Info, Arc/Info, GIS
- 104: ARC/INFO, ERDAS, ARC export files
- 108: ARC/INFO
- 110: MAP/INFO (I have encountered some boundary problems)
- 111: GIMMS, ARC/INFO, BNA format data output by DBD91 on MIDAS and UKBORDERS.
- 112: ARC/INFO IDRISI, MAP/INFO
- 119: ARC/INFO, ARCVIEW, MAPINFO
123. MAPINFO
- 132: ARC or any other comparable software
- 133: ARC/INFO, ARCVIEW
- 135: Use commercial formats but also an open non-proprietary format like the ESRC format for service developers.
- 136: MapInfo
- 137: WWW
- 138: On CD-ROM and UK BORDERS

4.11 Further comments

Questionnaire instructions:

Comments on the ESRC/JISC Census Programme

Further comments on the 2001 Census of Population

- 002: The 5 basic components are OK. But perhaps a dissemination programme should be added, in order to ensure that the best options are taken for user support and data dissemination, to which existing census programme units and others could apply.
- 007: Without a precise date for the release of 2001 data -- how can we plan?
- 011: The regionally-based workshops are particularly useful. Not always to get up and down to London for some people in the south-east region. More regional workshops please!
- 027: (1) It will be essential to continue to support the Census datasets in the next decade. Each data set will need storage, maintenance, dissemination, support, training, updating, development. (2) There should be an open bidding process for all elements of the programme though I would expect existing Units to be very strong contenders for data centre roles. The time is probably ripe for the creation of networked centres involving several linked institutions containing different expertises providing a collaborative service. (3) In the 1991-96 programme the Census Development Programme projects all worked to deliver their products to Census Data Units for use and dissemination. The Census Training Programme projects were less successful in this. In a 2001-06 programme all sub-programmes should have dissemination strategies linked to Census Data Units. (4) In 2001 Census Data Units will need to look after (a) Area Statistics; (b) Boundary Data; (c) Micro-data; (d) Interaction Data; (e) Longitudinal Data and (f) Flexible outputs.
- 047: I would suggest that the priority should be: (1) Data Centres; (2) Co-ordinator (with Census Offices); (3) R & D (i.e. research which enhances the Data); (4) Training. Seminars and other Research should be funded separately. If the Census can get 20,000 variables from its short census form, then this should produce a vast dataset - will it be accessible on-line?.
- 064: (1) Training now far less important than previously. (2) Attention should be focused on creating useful value added products of generic and general utility to census users: - Look up tables; - GB profiles; - USAR; etc. (3) There should be a census programme co-ordinator to front negotiations with ONS and oversee census programme.
- 065: In addition, it would be useful to have the option of getting copies of published report tables (100% and/or 10% statistics) in a format that can be input to LOTUS 123 or then spread sheet packages for analysis of census national statistics.
- 073: Surely there must be on going programmes supported by ESRC/JISC and ONS?
- 081: One-off training (i.e. other than MIDAS regular courses) have not made as much impact as hoped for. Leave that to happen, and fund central development that can parachute into anywhere, providing basic good training that will stimulate local training initiatives.
- 085: 1991 Programme seemed to work well.
- 088: (1) Training materials need to be kept up date over the 10 year period. (2) More exemplar based training materials. (3) Derived products (e.g. GBProfiles) need to be maintained.
- 098: Referring to 4 and 5: I would support the workshop recommendations esp. 1, 4, and 5. I find it difficult to say which software I might use - spreadsheets, databases, statistical packages/which can usually read spreadsheets e.g. Epi-info, Arcus, etc.).

- 132: I think the key issue is to encourage and support a “community” of Census users. This is important for exchange of information (and encouragement!) but also as a base on which to represent academic and subject specific interests to ONS etc.
- 133: All [aspects of the programme] have been important, useful and highly successful.
- 140: I think this has been successful, except for the research programme, the output from which that I have so far seen/refereed has been well below the standards I would expect from ordinary grants and predictably so from the original applications.