

AG (99) 17

CODING OCCUPATION AND INDUSTRY DATA

This paper discusses two issues in relation to coding of occupation and industry data for the 2001 Census:

§ Implementation of the revised Standard Occupation Classification (SOC 2000); and

Advisory Group members are asked to note the implications of the revision of the Standard Occupational Classification on occupation coding in the 2001 Census.

§ A revised filtering strategy for employment related information.

Advisory Group members are invited to give their views on the options suggested in this paper within two weeks from the date of the Advisory Group meeting.

Celia Moss
4200W
Data Needs
Census Division, ONS

IMPLEMENTATION OF THE REVISED STANDARD OCCUPATION CLASSIFICATION (SOC 2000)

1. Occupation coding on the 2001 Census

- 1.1 The Standard Occupational Classification was published in 1990 and used in the 1991 Census of Population. The classification is currently being revised and the next edition, to be called SOC 2000, will be published in Spring 2000.
- 1.2 The responses to the occupation questions on the 2001 Census will be coded to SOC 2000.

2. Background to revision of the Standard Occupational Classification

- 2.1 In 1997 the SOC Steering Group approved the scope of the revision and the main requirements of the work as follows:

- § The classification's principles and conventions should be reviewed but there should be **no** change to the fundamental principles of *skill level* and *nature of work activities* underlying the classification.
- § The classification should retain its basic structure, be up-to-date, but retain a reasonable degree of continuity with the existing SOC.
- § The contents of the unit groups should be assessed with the expectation that some of the existing unit groups will be divided and re-defined, others will be amalgamated.
- § The job title index should be updated with a view to improving the content for its use in automated coding, and for users with limited sources of data.
- § The revision should reflect recommendations and requirements from the work to create the National Statistics Socio-economic Classification.
- § The revision should attempt to improve the degree of harmonisation with ISCO 88(COM), where this can be achieved without compromising the other requirements.
- § Links back to current SOC unit groups should be tracked and a means to bridge back to them provided

- 2.2 One of the first decisions made on the structure of the classification was to bring the sub-major groups into the numbering system.

SOC 1990 Example of the numbering system2.1.1.1.1.1.1.1

2.1.1.1.1.1.1.2

Major group2Professional occupations2.1.1.1.1.1.1.3

Sub major group2aScience and engineering professionals2.1.1.1.1.1.1.4

Minor group20Science professionals2.1.1.1.1.1.1.5

Unit group200Chemists2.1.1.1.1.1.1.6

2.1.1.1.1.1.1.7

SOC 2000 (draft framework) Example of the numbering system 2.1.1.1.1.1.1.8

2.1.1.1.1.1.1.9

Major group2Professional occupations2.1.1.1.1.1.1.10

Sub major group21Science and technology professionals2.1.1.1.1.1.1.11

Minor group211Science professionals2.1.1.1.1.1.1.12

Unit group2111Chemists2.1.1.1.1.1.1.13

- 2.3 The pan European comparison of occupation counts, which indicated that the UK has too many managers, prompted a review of the definition of managers within SOC. This has resulted in the movement of jobs out of Major Group 1.

- 2.4 Other themes have featured in the research work associated with the revision of the classification. These include:

- § the latest developments in information and communications technology, such as the increased use of the internet;
- § the move away from common grades in the Civil Service;
- § the greater emphasis on qualifications and specialist skills required for those working in the protective services;
- § the use of telephone call centres for various selling and service activities.

3. SOC 2000

- 3.1 Whilst an objective of the SOC review has been to maintain a reasonable degree of continuity with the existing SOC, the main priority has been to bring the classification up-to-date, to reflect changes in society.
- 3.2 This has resulted in significant change to the classification. A considerable number of jobs have moved between unit groups, many between minor groups and more than a few between major groups. However the numbers of groups in the tiers are similar.

SOC 1990	SOC 2000 (draft framework) 2.1.1.1.1.1.1.14
9	Major groups 9Major groups 2.1.1.1.1.1.1.1.15
22	Sub major groups 25Sub major groups 2.1.1.1.1.1.1.1.16
77	Minor groups 81Minor groups 2.1.1.1.1.1.1.1.17
371	Unit groups 353Unit groups 2.1.1.1.1.1.1.1.18

- 3.3 Due to the number and nature of changes, there will be no direct ‘concordance’ produced between SOC90 and SOC 2000. However, mechanisms for bridging from SOC 1990 to SOC 2000 are being discussed within ONS in the context of key collections.
- 3.4 It is intended that the 2001 Census will be coded to SOC 2000. However, to assist users in comparing information from the previous Census with 2001 Census data, the Occupation Information Unit will produce a table showing the ‘relationship’ between the unit groups of SOC 1990 and SOC 2000. As part of the SOC revision process the records in the 1991 Census Economic Activity Subsample are being re-classified into the new unit groups, providing a source of the relationship information. We have yet to finalise the method of presentation, but it will probably be a table providing percentages by sex as illustrated below. This information will be available prior to the release of data from the 2001 Census.

Relationship between SOC 1990 and SOC 2000 unit groups - Example of presentation style (percentages for illustration only)

SOC 1990	SOC 2000 2.1.1.1.1.1.1.1.19
Unit group	Percentage Unit group 2.1.1.1.1.1.1.1.20
	Males Females 2.1.1.1.1.1.1.1.21
360	Estimators, valuers 100.0100.03531 Estimators, valuers and assessors 2.1.1.1.1.1.1.1.22
361	Underwriters, claims assessors, brokers, investment analysts 20.010.03531 Estimators, valuers and assessors 2.1.1.1.1.1.1.1.23
	30.015.03532 Brokers 2.1.1.1.1.1.1.1.24
	25.05.03533 Insurance underwriters 2.1.1.1.1.1.1.1.25
	25.070.03534 Finance and investment analysts/advisors 2.1.1.1.1.1.1.1.26

4. Action

- 4.1 Advisory Group members are asked to note the implications of the revision of the Standard Occupational Classification on occupation coding in the 2001 Census.

A REVISED FILTERING STRATEGY FOR EMPLOYMENT-RELATED INFORMATION

1. Introduction

- 1.1 At the June 1998 UKCC meeting it was agreed that further consideration would be given to the age-cut off for collecting employment-related information from the 2001 Census. Since that meeting, further consideration has been given to the volume of responses likely to require processing using the Census Rehearsal filtering strategy, the level of quality of these responses, and the burden on the public.
- 1.2 This paper puts forward revised proposals in relation to the collection and processing of responses to employment related questions in the 2001 Census. It looks at the background surrounding proposals for the occupation and industry questions in particular. Issues relating to quality and costs are raised, and a number of options put forward.

2. Background

1991 Census

- 2.1 In the 1991 Census, employment-related questions were answered by all persons aged 16 or over who were either currently working, or who had a paid job within the last 10 years. For occupation and industry questions, only a 10% sample of responses were coded.

Consultation and research on user requirements

- 2.2 Extensive consultation and research has been carried out on the questions proposed for inclusion in the 2001 Census. In particular, the Labour Market Subgroup of the Content Working Group has considered issues surrounding the employment-related questions.
- 2.3 The main user requirements for this information are described below.
 - § 100% coding - There has been a strong user requirement for the 2001 Census that occupation and industry be coded for 100% of responses.
 - § Information on people who have ever worked - A number of users expressed a requirement for extending the coding of information for people not currently working to all people who had ever worked for the following purposes.
 - Users of the National Statistics Socio-economic Classification (NS-SEC) require that people who have retired or have not worked for some time (e.g. if they are looking after the home and family) to be classified on the same basis as those currently working.

- Occupational health analysis. The information is used when looking at which occupations and industries certain age groups (in particular the elderly) were employed in, to analyse and anticipate levels of present and future health problems related to those particular occupations and industries.

- To analyse patterns of unemployment for particular occupations and industries.

§ ‘Open’ occupation and industry questions – Due to the high cost of coding occupation and industry data, ‘closed’ questions (which provide a limited list of pre-coded tick-box responses) were tested in 1997/1998 using the ONS Omnibus Survey. There was considerable resistance from various users, in particular the Dept. of Health, Department for Education and Employment and the ONS (Demography and Health) to this approach, particularly with respect to the occupation question. The Census Offices concluded that closed occupation and industry questions could not effectively meet user needs.

2.4 Small scale testing using the ONS Omnibus Survey demonstrated that retired respondents objected to answering questions that were no longer relevant or appropriate. Therefore, as well as the minimum age cut-off for employment-related questions, an additional age cut-off of 75 and over was recommended.

1999 Census Rehearsal

2.5 In order to meet user requirements, the 1999 Census Rehearsal applied an age cut-off of under 16, and 75 years and over, with no cut-off for those not currently working. The following chart reflects the filtering strategy applied to employment related questions in the Census Rehearsal.

Questions	Respondents filtered out
<i>Age Filter</i>	<u>Under 16's, 75 and over</u>
	Qualification Level, Professional Qualifications, Activity Last Week
<i>Ever-worked filter</i>	<u>Never worked</u> (inc. year in which last worked)
	Employment Status, Size of Organisation, Occupation, Supervisor status, Industry
<i>Currently Working Filter</i>	<u>Not currently working</u>
	Name of Employer, Address of Workplace, Means of Travel to Work, Hours Worked

3. Issues for consideration

- 3.1 User requirements for information must be balanced against achieving quality output, and ensuring the best possible value from the investment in the Census. Consideration must also be given to public acceptability and the possible burden on the public of collecting required information.
- 3.2 Automatic and computer assisted coding of write-in responses to the occupation and industry questions have meant that 100% of responses can be processed. However, the cost of processing these questions is still high due to the complexity of responses and the likely level of quality. Coding costs increase significantly if manual coding is required.
- 3.3 There is evidence to suggest that the quality of responses to occupation and industry questions diminishes as the respondents' age increases. This can be demonstrated by the rise in the percentages of people in the occupation category '*not stated or inadequately described*' in 1991 outputs. The number of '*not stated or inadequately described*' was approximately 1% for the 45-54 year age group, 3% for the 70-74 year age groups, and 6% for 75 and over age group. Of those in the older age groups who were allocated a 'valid' occupation in 1991, a higher percentage of responses had to be 'expert' coded, using additional sources of information.
- 3.4 There is also some evidence from testing that the quality of responses also decreases as the time since the respondent last worked increases. Thus the cost of coding those people who are not currently working is likely to be disproportionately high compared to those who are in employment.
- 3.5 The high number of calls to the Census Rehearsal Helpline in relation to the employment questions confirms concerns in relation to the relevance and acceptability of these questions to older respondents and to those who last worked some time ago. Whilst some respondents had problems in deciding and remembering what to record, others took issue about the relevance of the information that was being asked for. A selection of comments from the Census Helpline illustrate this:

" Had not worked for the past 15 years and has been a housewife since. Felt the questions were not relevant"

" Main employment section is difficult to answer. Last worked 30 years ago. How is she to remember all the details?"

"Caller finds it ridiculous that he should answer questions on employment when he is retired"

4. Options for the 2001 Census

- 4.1 The Census Offices propose that for the 2001 Census 100% of responses to the occupation and industry questions will be coded. In addition, to provide the level of detail required, open questions will be asked. However, in recognition of issues in relation to quality, the high cost of processing, the level of acceptability and the burden on the public, the Census Offices cannot justify the collection and processing of employment information for all those respondents included in the Census Rehearsal.
- 4.2 There are three ways to address the issues outlined in Section 3 (in terms of quality, cost and public acceptability). The first method is to reduce the age cut-off; the second is to introduce a filter based on the number of years since last worked (as in 1991); and the third method is to use a combination of both age and years since last worked. Using recent data, Annex A shows the effects of these filters set at varying levels.
- 4.4 This paper puts forward 3 options from the many permutations shown in Annex A. Note that the cost of processing information on occupation and industry beyond those currently in employment (approximately 27.7 million responses) has been calculated at between £250,000-£300,000 per 1 million additional responses.

Option 1 - Code only those people aged 16 to 74 who are currently working.

This option would provide employment details as they exist at the time of the Census, providing a snapshot of the UK at that time, rather than capturing employment information relevant X number of years ago. The information would be of more consistent quality, would seem more relevant to respondents thus improving acceptability and reducing the burden on the public, and would be significantly cheaper to process than Options 2 and 3.

Option 2 - Code all people aged 16 to 69 years who are currently working or have worked in the past 5 years.

This option would cost £1 million (approx.) more than Option 1. Additional information would be collected for people who were not currently employed, however fewer responses would be collected from older people. As older respondents question the relevance of providing this information, this option would improve acceptability. In addition, a cut-off for years since last worked would improve the overall quality of responses compared to the approach taken in the Census Rehearsal.

Option 3 - Code all people aged 16 to 74 years who are currently working or have worked in the last 10 years.

This option would cost an additional £2.5 million (approx.). It would retain comparability with 1991 in terms of the cut-off for those not currently working. In terms of the age cut-off, information would be collected on the same basis as the Census Rehearsal. Issues in relation to the burden on older people are not addressed by this option. However, it does address issues in relation to the relevance of the question to those who last worked more than 10 years ago. In addition, a cut-off for years since last worked would improve the overall quality of responses compared to the approach taken in the Census Rehearsal, although to a lesser extent than Option 2.

5. Action

- 5.1 Additional costs to collect and process this information need to be justified on the basis of a strong user requirement, and must be weighed up against alternative uses to which resources could be put. **Advisory Group members are invited to give their views on the options suggested in this paper within two weeks from the date of the Advisory Group meeting. When responding, Advisory Group members are asked to provide clear justification for any additional expenditure.**

Estimated numbers of people requiring coding using combinations of age and time since last worked filters

figures in millions

UK figures	Age cut off	2.1.1.1.1.1.1.1.27
Time since last employed	60 years Pension –able age	65 years 70 years 75 years
	2.1.1.1.1.1.1.1.28	
		2.1.1.1.1.1.1.1.29
Working		25.926.727.027.327.7#2.1.1.1.1.1.1.1.30
		2.1.1.1.1.1.1.1.31
Working, or,		2.1.1.1.1.1.1.1.32
Worked in last year		27.528.428.829.229.62.1.1.1.1.1.1.1.33
Worked in last 2 years		28.329.229.730.230.72.1.1.1.1.1.1.1.34
Worked in last 3 years		28.929.930.531.131.62.1.1.1.1.1.1.1.35
Worked in last 4 years		29.330.431.031.832.32.1.1.1.1.1.1.1.36
Worked in last 5 years		29.730.831.532.5+33.02.1.1.1.1.1.1.1.37
Worked in last 6 years		30.131.232.133.233.72.1.1.1.1.1.1.1.38
Worked in last 7 years		30.431.632.633.834.42.1.1.1.1.1.1.1.39
Worked in last 8 years		30.732.033.034.435.12.1.1.1.1.1.1.1.40
Worked in last 9 years		31.032.433.434.935.72.1.1.1.1.1.1.1.41
Worked in last 10 years		31.332.733.835.536.4*2.1.1.1.1.1.1.1.42
Ever worked		32.634.135.438.040.62.1.1.1.1.1.1.1.43
		2.1.1.1.1.1.1.1.44
Never worked		1.91.92.02.02.02.1.1.1.1.1.1.1.45
		2.1.1.1.1.1.1.1.46
Total		34.536.037.440.042.62.1.1.1.1.1.1.1.47

* Option 1
+ Option 2
Option 3

Sources of data

Mid year population estimates, ONS
December 1998 Labour Market Statistics, ONS
Labour Force Survey (commissioned outputs), ONS
Estimates based on these official figures