Office for National Statistics

CENSUS ADVISORY GROUP

AG (10) 04 EILR Classifications

Coding classifications for the 2011 Census

1. Introduction

The classifications to be used when coding data from the 2011 Census have now been finalised, and are the largest ever developed for a UK Census. They have been developed by business areas across ONS, with some being developed specifically for the 2011 Census and others being standard ONS classifications or slight, census specific, variations of these.

The classifications are used to code responses to a number of questions, following the data capture operation, and prior to ONS receiving the data. The coded data is then used in output production, and outputs cannot easily be produced for responses that do not have an entry on the classifications[1].

This paper outlines the different classifications to be used when coding data from the 2011 Census, and the process used to develop them.

Advisory Group members are asked to note the work undertaken in the development of these classifications.

2. Classifications required

Coding classifications are required when coding a range of topics in the 2011 Census – specifically, those that have a "write in" response option and, therefore, a wide range of possible responses.

The topics for which a classification is required for the 2011 Census in England and Wales are:

- Country of birth
- Country of non-UK address (in second residence, address one year ago and visitor questions[2])
- Passports held
- National Identity
- Ethnic Group
- Religion

---

[1] Written responses that do not map to the classifications will be assigned to 'uncodeable' and will be imputed downstream. However, as the captured text will also be on the data delivered to ONS, it may be possible to produce non-standard outputs for a specific category if necessary.
[2] Also for workplace address in Northern Ireland.

- Language
- Occupation
- Industry
- National Statistics Socio-Economic Classification (NS-SEC)[3]

Although there are ten topics which require a classification for coding, the first two of these can use the same (country) classification.

2.1 Country classification

The classification to be used for country of birth, and country of non-UK addresses, is a standard ONS product, with a slight census specific variation.

This classification is a list of all UK-recognised countries, using their current names, and is maintained and updated by ONS's Classifications and Harmonisation Unit (CHU). A code for 'Island of Ireland (not otherwise specified)' has been added to cover situations of respondents ticking both 'Northern Ireland' and 'Ireland'.

The country classification has also been modified to provide a classification for passports held.

2.2 National identity, Ethnic group and Religion classifications

The classifications to be used when coding National Identity, Ethnic Group and Religion in the 2011 Census are based on the standard ONS classifications for these topics, and have been developed by ONS's Ethnicity and Identity Branch (EIB).

These classifications are comparable with those used in the 2001 Census, with the exception of National Identity which is new, and have been expanded following consultation with a range of users and stakeholders by EIB in 2009.

These classifications contain an element of overlap, where it is possible for respondents to give the same response to more than one question (for example Welsh national identity and ethnicity, or Sikh ethnicity and religion).

The coding rules that apply the ethnic group classification have also been improved following user consultation, to ensure that rare responses can be correctly aggregated into the major categories for outputs.

2.3 Language classification

Prior to the 2009 Census Rehearsal, ONS did not have a language classification. Therefore, the Census Statistical Design team developed one, based on existing classifications used in academia and by National Statistics institutes worldwide.

This classification was used to code responses to the 2009 Census Rehearsal, and was also included in the consultation run by EIB discussed in section 2.2. Following a review of the coding of the Rehearsal, and responses to the consultation, a range of codes were removed (the initial classification contained a small number of duplicates), reclassified (for aggregation) or added.

---

[3] NS-SEC is not applied prior to ONS receiving the data, but is applied downstream by ONS.

2.4 Occupation and Industry classifications

The classifications for Occupation and Industry will be a census-specific variation of the current ONS Standard Occupation and Industry classifications (SOC 2010 and SIC 2007).

The main classifications and the census-specific versions were developed by the CHU in ONS, who consulted widely with the user community when doing so.

These classifications are broadly comparable with the earlier versions used in previous censuses, and will now be used for the coding of these variables in all ONS surveys.

2.5 NS-SEC

The NS-SEC will be updated for use in the 2011 Census late in 2010, and will be comparable with that used in 2001.

NS-SEC is derived based on responses to a range of questions but, because the 2011 Census will not include a question on 'number of employees at workplace', the *Reduced* derivation method will be used, rather than the *Full* method used in 2001. Previous analysis has estimated that this will only have a small impact on derivation accuracy and that, given the inconsistency of responses to this question in the 2001 Census, and the high level of item non-response experienced, it will not affect the utility of these data.

3. UK Harmonisation

As with all aspects of census development for 2011, ONS have worked closely with colleagues in the General Register Office for Scotland (GROS) and the Northern Ireland Statistics and Research Agency (NISRA) to ensure consistency between all the UK Censuses where possible. The classifications developed by ONS will also be used in Scotland and Northern Ireland.

4. Use in coding

Although the classifications are used to code the census responses, they are not sufficient for this purpose without a more comprehensive reference, called a Coding Index.

The classification contains all the codes required for outputs, but the index expands on this by including a range of possible responses for each code. These responses include common mis-spellings and previous/alternative names for codes (for example, specifying that 'Burma' and 'Myanmar' be given the same code in the Country of Birth classification).

The coding indices are developed by ONS Census division, in consultation with topic experts across ONS and based on written responses to the 2001 Census, 2007 Census Test and 2009 Census Rehearsal.

Where a response is given that is not on the classification or the index, it is coded as 'uncodeable' and a response will be imputed. The text response is also included on the output database which will inform future classification and index development.

5. Comparability with 2001

The classifications for passports held, national identity, and language are new for the 2011 Census and therefore do not cause any issues for comparability with 2001.

SOC, SIC, and NS-SEC are all standard classifications, and whilst there are differences between the versions used in the 2001 Census and those that will be used in 2011, the classifications are broadly comparable and documentation which enables comparisons between the different versions of the classifications to be made is produced by CHU.

The ethnic group and religion classifications have been expanded since 2001 to include new categories and provide a greater level of detail. Work is currently ongoing to identify and address any comparability issues.

6. Outputs

All relevant outputs from the 2011 Census will be built using the above classifications. The recent consultation, on the design of 2011 Census outputs, used these classifications in the development of a baseline specification, and will be used again in the autumn where a revised 'user defined' set of specifications will be issued. Users will therefore have a second opportunity to consider the classifications developed when responding to consultation about the outputs required, as outputs will not be possible if a required category does not have a specific entry on the classification.