

Northern Ireland Census 2011



2011 Census Microdata

Sample of Anonymised Records – Teaching File

Product Overview

Version 2.0

January 2014

Updated February 2015



Contents

1	Introduction.....	2
2	What are Microdata SARs	4
3	Preserving the confidentiality of the data	4
4	2011 Census Microdata SARs Teaching File	5
5	Purpose of the 2011 Census Microdata SARs Teaching File	7
6	Next Steps	8
7	How the Teaching File should be cited	9
	Appendix 1: Contents of the Teaching File.....	10

1 Introduction

- 1.1 In terms of scale, the Census of Population and Housing is without doubt the largest and most complex statistical exercise undertaken in Northern Ireland. Traditionally, it has been conducted every ten years and requires individuals living in either a household or a communal establishment to provide answers to a variety of questions spanning different topics of interest. The 2011 Census Household Questionnaire¹ details the questions and topics that were included in the most recent Census, which was conducted on 27 March 2011 in line with arrangements across the rest of the United Kingdom (UK). It can be seen that while the vast majority of questions could be answered by simply ticking the appropriate box/boxes, provision was made for write-in responses where appropriate.
- 1.2 The Census is renowned and valued for providing comprehensive and robust information that is comparable and consistent for both small areas and small population groups. Such information has many important applications and is extensively used across the public, private and voluntary sectors. In addition it is widely used in academia, largely for teaching and research purposes.

¹ The 2011 Census household questionnaire is available from:
<http://www.nisra.gov.uk/archive/census/2011/forms/household.pdf>

- 1.3 The vast majority of the information released from the 2011 Census in Northern Ireland has been produced in tabular form and released through the Northern Ireland Neighbourhood Information Service (NINIS) website.² Such information is particularly helpful in understanding the characteristics of, for example, households and people living in a particular geographical area and/or people in a particular population sub-group. In order to complement these statistical aggregates, Census Office is developing Microdata samples of anonymised Census records (SARs). The purpose of this paper is to provide an overview of the first of the planned 2011 Census Microdata SARs products (i.e. the 2011 Census Microdata Teaching File³), which was released by NISRA on 23 January 2014.
- 1.4 In releasing this product, NISRA both acknowledges and very much values the important work undertaken by the UK Data Service⁴ in terms of both promoting and supporting the use of all Microdata SARs products released from the 2011 and earlier Censuses.

² Census 2011 section of the NINIS website:

<http://www.ninis2.nisra.gov.uk/public/Theme.aspx?themeNumber=136&themeName=Census%202011>

³ The 2011 Census Microdata Teaching File is available to download from the NINIS website:
http://www.ninis2.nisra.gov.uk/Download/Census%202011_Winzip/2011/Microdata%20File.zip

⁴ Further information on the UK Data Service is available from:
<http://ukdataservice.ac.uk/>

2 What are Microdata SARs

2.1 SARs, as the name suggests, are samples of individual records drawn from the Census outputs database which have been completely anonymised. Such samples can be based on individual person level records or individual household level records and have typically been utilised for research purposes. Steps are taken to preserve the confidentiality of all individual records (both person based and household based) that are included in any such sample. In addition, access arrangements are very much tailored to reflect both the nature and detail of the information included for each record.

3 Preserving the confidentiality of the data

3.1 The information gathered through the Census is captured and processed electronically. Such processes include, for example, (i) the coding of the responses provided to each question and topic (both for tick box responses and write-in responses) and (ii) the imputation of missing details (both for people who didn't respond to the Census and for people who did respond but omitted to answer particular questions). Following this processing, the resulting information is collated into appropriate categories for release through the various planned output products.

3.2 In releasing any information from the Census, steps are taken to ensure that the confidentiality of the information provided by the general public in their questionnaire is completely preserved. This is achieved through the application of Statistical Disclosure Control (SDC) protocols, which are common across the UK and are appropriately tailored to reflect the content and access arrangements associated with the various output products.

3.3 In addition to the various SDC measures that are deployed more generally throughout the processing of the information gathered through the Census (e.g. prior to the publication of statistical aggregates), specific measures are taken in respect of the SARs products to ensure that they are completely anonymised. Without revealing the full range of measures that are deployed to afford the SARs the appropriate level of protection (which is paramount and part of the protection mechanism), it is important to stress that:

- all personal identifiers (e.g. names, addresses and dates of birth) have been completely removed from the records included in the sample; and
- potentially disclosive output variables have either been completely removed from the sample records or the output categories provided for such variables have been suitably collapsed in order to preserve confidentiality.

4 2011 Census Microdata SARs Teaching File

4.1 The main population base for published statistical tables from the 2011 Census in Northern Ireland is the usual resident population base as at Census day, 27 March 2011. By way of background, for 2011 Census purposes a usual resident of the United Kingdom (UK) is anyone who, on Census day, was in the UK and had stayed or intended to stay in the UK for a period of 12 months or more, or had a permanent UK address and was outside the UK and had intended to be outside the UK for less than 12 months.

- 4.2 Against this background, the 2011 Census Microdata SARs teaching file comprises a sample of 19,862 records (approximately 1 per cent) relating to people who were usually resident in Northern Ireland at the time of the 2011 Census. For each individual, information is available for seventeen separate characteristics (e.g. sex, age, marital status) to varying degrees of detail. Both the size of the sample and the content of the records in the file have been harmonised, wherever possible, with the equivalent SARs teaching file that the ONS simultaneously released for England and Wales.
- 4.3 This is the first such sample of Census records to be accessible to the general public. The file, which has been released as a comma separated format text file (.csv), can be accessed through the Northern Ireland Neighbourhood Information Service (NINIS).³
- 4.4 Appendix 1 details the variables along with the associated output categories and value labels that have been included for each record in the teaching file.
- 4.5 The distributions of each variable within the teaching file have been compared with the distributions from the 2011 Census to show that the teaching file is representative of the 2011 Census. While this work demonstrates the representative nature of the teaching file, it should be noted that the teaching file is subject to sampling error. Accordingly any inferences that are drawn need to be considered in the context of the more detailed tabular data that has already been released from the 2011 Census and/or the findings that emerge from the more detailed Microdata SARs products that are planned.

5 Purpose of the 2011 Census Microdata SARs Teaching File

5.1 As with previous Censuses, it is envisaged that a variety of Microdata SARs products will be released from the 2011 Censuses across the UK. To this end, the Census Offices are continuing to work alongside key users to develop and agree the specifications and access arrangements for those products, with the aim of further releases later in 2014.

5.2 Ahead of this, the primary purpose of the teaching file, which comprises unit-record level data as opposed to statistical aggregates, is as an educational tool aimed at:

- encouraging wider use of Census data by facilitating another way of examining Census data, for example through the building of statistical models, over and above that already available through the raft of standard tabular output released to date;
- providing a broad insight into the sort of detail that is generally included in a SARs product, along with data formats and any associated metadata. This will enable users (arguably those less experienced at using SARs products) to ‘play’ with the data and increase their knowledge and skills in readiness for accessing the more detailed SARs products that are planned and will be available in, for example, a safe setting; and
- assisting with the teaching of statistics and geography at GCSE and higher levels.

6 Next Steps

- 6.1 As outlined above, the Census Offices are continuing to work alongside key users to develop and agree the specifications and access arrangements for more detailed 2011 Census Microdata SARs products. The broad direction of travel is towards the provision of larger samples than was the case in 2001 and to make the products as widely accessible as possible, within the constraints imposed by the detail of their content.
- 6.2 Users are being kept informed about developments and timescales through the release plans section of the Northern Ireland 2011 Census Output Prospectus⁵ and the UK Data Service⁴. Current planning is on the basis that further products will be released later in 2014, when the specifications and sample data have been assessed from a disclosure perspective.
- 6.3 At the time of writing, two further Individual person based SARs products and one individual household based SARs product are currently being considered, as detailed below.
- **Secure Products**: An anonymised sample of individuals and an anonymised sample of households – the maximum size of these samples will be 10%. These samples will be non-overlapping and will be the most detailed available. Accordingly, they will only be available in the Secure Setting of the Virtual Microdata Laboratory (VML) managed by the Office for National Statistics (ONS). It is envisaged that users in Northern Ireland

⁵ 'Northern Ireland 2011 Census Output Prospectus' is available from:
<http://www.nisra.gov.uk/Census/2011CensusProposedOutputs.html>

will be able to access these files from within the secure setting in NISRA Headquarters using the secure link to the VML.

- **Safeguarded Products**: An anonymised sample of individuals, which will be available to approved researchers via specific end-user licence arrangements – the maximum size of the sample will be 5%. This sample will be drawn from, and hence overlap with, the proposed secure sample (maximum of 10%) of individuals mentioned above. The level of detail included for each record in the sample will however be adjusted (i.e. it will be less detailed) to reflect the associated access arrangements and the need to preserve the confidentiality of the information, which is paramount.

7 How the Teaching File should be cited

7.1 The Open Government Licence⁶, which applies to the 2011 Census Microdata SARs teaching file, allows unrestricted use of Government data as long as the source is acknowledged. This can be achieved by including the following when publishing any analyses that have been based on the teaching file.

"These statistics may be used, excluding logos, under the terms of the Open Government Licence."

⁶ Further information on the Open Government Licence is available from:
<http://www.nationalarchives.gov.uk/doc/open-government-licence/version/>

Appendix 1: Contents of the Teaching File

The teaching file is available as a comma separated format text file (.csv) from the NINIS website:

http://www.ninis2.nisra.gov.uk/Download/Census%202011_Winzip/2011/Microdata%20File.zip

Variable Name	Description	Number of categories	Variable Values	Variable Labels
RECORD_NUMBER	Record number	-	-	Unique reference ID
REGION	Region	10	N92000002	Northern Ireland
RESIDENCE_TYPE	Residence Type	2	C H	Resident in a communal establishment Not resident in a communal establishment
FAMTYPE	Family composition	6 (also -9)	1 2 3 4 5 6 -9	Not in a family Married/same-sex civil partnership couple family Cohabiting couple family Lone parent family (male head) Lone parent family (female head) Other related family No code required (Resident of a communal establishment)
POPBASE	Population base	1	1	Usual resident
SEX	Sex	2	1 2	Male Female
AGE	Age	8	1 2 3 4 5 6 7 8	0 to 15 16 to 24 25 to 34 35 to 44 45 to 54 55 to 64 65 to 74 75 and over

Variable Name	Description	Number of categories	Variable Values	Variable Labels
MARSTAT	Marital Status	5	1 2 3 4 5	Single (never married or never registered a same-sex civil partnership) Married or in a registered same-sex civil partnership Separated but still legally married or separated but still legally in a same-sex civil partnership Divorced or formerly in a same-sex civil partnership which is now legally dissolved Widowed or surviving partner from a same-sex civil partnership
STUDENT	Schoolchild or student in full-time education	2	1 2	Yes No
COB	Country of Birth	2	1 2	UK Non UK
HEALTH	Health	5	1 2 3 4 5	Very good health Good health Fair health Bad health Very bad health
ETHNICITY	Ethnic group	2	1 2	White Other ethnic group
RELIGIONNI	Religion belonged to	8	1 2 3 4 5 6 7 8	Roman Catholic Presbyterian Church in Ireland Church of Ireland Methodist Church of Ireland Other Christian (including Christian related) Other religions and philosophies No religion Religion not stated
RELIGIONBNI	Religion belonged to or brought up in	4	1 2 3 4	Catholic Protestant and other Christian (including Christian related) Other religions None

Variable Name	Description	Number of categories	Variable Values	Variable Labels
ECONACT	Economic Activity	9 (also -9)	1 2 3 4 5 6 7 8 9 -9	Economically active: Employee Economically active: Self-employed Economically active: Unemployed Economically active: Full-time student Economically inactive: Retired Economically inactive: Student Economically inactive: Looking after home or family Economically inactive: Long-term sick or disabled Economically inactive: Other No code required (Aged under 16)
OCC	Occupation	9 (also -9)	1 2 3 4 5 6 7 8 9 -9	Managers, Directors and Senior Officials Professional Occupations Associate Professional and Technical Occupations Administrative and Secretarial Occupations Skilled Trades Occupations Caring, Leisure and Other Service Occupations Sales and Customer Service Occupations Process, Plant and Machine Operatives Elementary Occupations No code required (Aged under 16, people who have never worked)

Variable Name	Description	Number of categories	Variable Values	Variable Labels
INDUSTRY	Industry	12 (also -9)	1 2 3 4 5 6 7 8 9 10 11 12 -9	Agriculture, forestry and fishing Mining and quarrying; Manufacturing; Electricity, gas, steam and air conditioning system; Water supply Construction Wholesale and retail trade; Repair of motor vehicles and motorcycles Accommodation and food service activities Transport and storage; Information and communication Financial and insurance activities; Intermediation Real estate activities; Professional, scientific and technical activities; Administrative and support service activities Public administration and defence; compulsory social security Education Human health and social work activities Other community, social and personal service activities; Private households employing domestic staff; Extra-territorial organisations and bodies No code required (Aged under 16, people who have never worked)
HOURS	Hours worked per week	4 (also -9)	1 2 3 4 -9	Part-time: 15 or less hours worked Part-time: 16 to 30 hours worked Full-time: 31 to 48 hours worked Full-time: 49 or more hours worked No code required (Aged under 16, people not working)