

census
2021

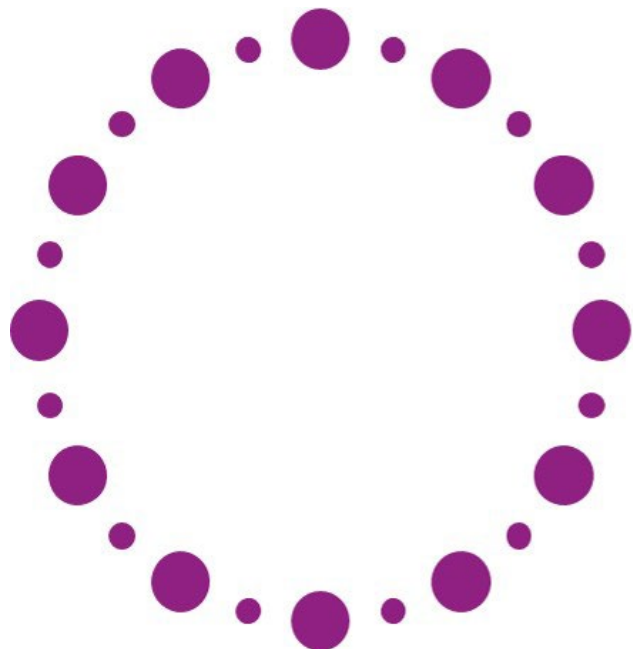


Microdata Sample of Anonymised Records

Public microdata teaching sample for Northern Ireland

Product overview

January 2025 – Version 1.0



nisra.gov.uk/census2021

Contents

1. Introduction	3
2. What are Microdata SARs.....	4
3. Preserving the confidentiality of the data	4
4. Census 2021 public microdata teaching sample	5
5. Purpose of the Census 2021 public microdata teaching sample.....	6
6. Next steps	7
7. How the Census 2021 public microdata teaching sample should be cited	8
8. Contact details	8
9. Accredited Official Statistics	8

1. Introduction

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 [Census 2021 Household Questionnaire](#) details the questions and topics that were included in the most recent census, which was conducted on 21 March 2021. 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.

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.

The vast majority of the information released from the Census 2021 in Northern Ireland has been produced in tabular form and released through the [Census 2021 Flexible Table Builder](#), the [Census 2021 Main Statistics](#) and [Census 2021 public microdata teaching sample](#) webpages. 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 Census 2021 Microdata SARs products i.e. the [Census 2021 public microdata teaching sample for Northern Ireland](#), which was released by NISRA on 16 January 2025.

In releasing this product, NISRA both acknowledges and very much values the important work undertaken by the [UK Data Service](#), in terms of both promoting and supporting the use of all Microdata SARs products released from the 2021 and earlier censuses.

2. What are Microdata SARs

SARs, as the name suggests, are samples of individual records which are 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

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.

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.

In addition to the various SDC measures that are deployed more generally throughout the processing of the information gathered through the census (for example, 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 (for example, 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. Census 2021 public microdata teaching sample

The main population base for published statistical tables from the 2021 Census in Northern Ireland is the usual resident population base as at census day, 21 March 2021. By way of background, for Census 2021 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.

Against this background, the Census 2021 public microdata teaching sample comprises a sample of 16,894 records (approximately 1 per cent) relating to people who were usually resident in Northern Ireland at the time of the 2021 Census. For each individual, information is available for sixteen separate characteristics (for example, sex, age, marital status) to varying degrees of detail. Both the size of the sample and the content of the records in the public microdata teaching sample have been harmonised, wherever possible, with the equivalent SARs teaching sample that the ONS released for England and Wales.

This is the first such sample of Census 2021 records to be accessible to the general public. The public microdata teaching sample, which has been released as a comma separated format text file (.csv), can be accessed through the [Census 2021 public microdata teaching sample for Northern Ireland](#) webpage.

The accompanying [Census 2021 public microdata teaching sample data dictionary](#) details the variables along with the associated output categories and value labels that have been included for each record in the public microdata teaching sample.

The distributions of each variable within the public microdata teaching sample have been compared with the distributions from the 2021 Census to show that the public microdata teaching sample is representative of the full 2021 Census. While this work demonstrates the representative nature of the public microdata teaching sample, it should be noted that the public microdata teaching sample 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 Census 2021 and/or the findings that emerge from the more detailed Microdata SARs products that are planned.

5. Purpose of the Census 2021 public microdata teaching sample

As with previous censuses, it is envisaged that a variety of Microdata SARs products will be released from the 2021 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 2025.

Ahead of this, the primary purpose of the public microdata teaching sample, 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

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 Census 2021 Microdata SARs products.

Users are being kept informed about developments and timescales through the release plans section of the [Census 2021 output prospectus](#) and the UK Data Service. Current planning is on the basis that further products will be released later in 2025, when the specifications and sample data have been assessed from a disclosure perspective.

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 Trusted Research Environment (TRE) Settings managed by the Office for National Statistics (ONS) and NISRA.
- **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 Census 2021 public microdata teaching sample should be cited

The [Open Government Licence](#), which applies to the Census 2021 public microdata teaching sample, 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 Census 2021 public microdata teaching sample. "These statistics may be used, excluding logos, under the terms of the Open Government Licence."

8. Contact details

If you have any queries regarding microdata products for Northern Ireland, you can contact Census Customer Services:

Email: census@nisra.gov.uk

Telephone: 028 9025 5156

9. Accredited Official Statistics

These are accredited official statistics. They have been independently reviewed by the Office for Statistics Regulation (OSR) and found to comply with the standards of trustworthiness, quality and value in the [Code of Practice for Statistics](#). This broadly means that the statistics:

- meet user needs
- are presented clearly and accessibly
- are produced using appropriate data and sound methods
- are managed impartially and objectively in the public interest