

Census 2021 Comprehensive Microdata (C21CM) User Guide (v1)

October 2024

C21CM Data: A Guide for Users

Contents

C21CM Data: A Guide for Users	2
1.0 Introduction to the C21CM Guidance Document	3
2.0 The C21CM Research Database	4
2.1 Overview	4
3.0 Data Collection Timeline and Years of Data Available	4
3.1 Data Collection Timeline	4
4.0 Data Quality Assurance	5
4.1 Quality Assurance of the C21CM Research Database	5
4.1.1 Consistency within the data	5
4.1.2 Consistency with published reports	5
5.0 Accessing the C21CM	5
5.1 Accessing the C21CM for research	5
5.2 Restricted Variables	6
6.0 Acknowledgement	7

1.0 Introduction to the C21CM Guidance Document

This document provides guidance to researchers on the Census 2021 Comprehensive Microdata (C21CM) database. The C21CM is a standalone, de-identified database comprised of 98.5% of Census 2021 person and household data for the population of Northern Ireland.

Included in this document are background information on the data source of the C21CM, a description of the C21CM and how researchers can access the data. When applying to use the C21CM data this document should be used in conjunction with the C21CM variable list.

All the uses of the C21CM come under Part 5, Chapter 5, Section 64 of the Digital Economy Act (2017).

The owner of this document is Northern Ireland Statistics and Research Agency – Research Support Unit (NISRA-RSU), and any changes to the document should be suggested to NISRA-RSU who will make amendments if required.

2.0 The C21CM Research Database

2.1 Overview

The census of Northern Ireland is undertaken by the Northern Ireland Statistics and Research Agency (NISRA). It usually takes place once every ten years and is the largest and most complex statistical exercise undertaken here. Census statistics are a vital source of information and are widely used by government, public bodies, academia, commercial businesses and others to develop policies, allocate resources and help deliver services. The last census in Northern Ireland was taken on 21 March 2021

The C21CM is a large microdata sample that covers 98.5% of records from Census 2021. This is the first time such a product has been made available in this way for research purposes in Northern Ireland.

3.0 Data Collection Timeline and Years of Data Available

3.1 Data Collection Timeline

- The Northern Ireland Census 2021 collected information on the usually resident population of Northern Ireland on Census Day (21 March 2021).

4.0 Data Quality Assurance

4.1 Quality Assurance of the C21CM Research Database

The RSU began data cleansing after receipt of the datasets. This was carried out separately on each of the datasets. The aim of this task was to ensure the accuracy and reliability of the data. RSU used the figures published in the [Census Flexible Table Builder](#) for comparison purposes. RSU applied a series of quality assurance checks to the data. These checks fell under the following headings:

4.1.1 Consistency within the data

RSU checked the data for anomalies, unexpected values, and issues with aggregations. Any resulting queries were relayed to the data providers for further information and advice to resolve the issues.

4.1.2 Consistency with published reports

RSU sought to replicate the key Census 2021 results published on the [Census Flexible Table Builder](#) Census 2021 tables.

The C21CM dataset is comparable to the key results published on the Flexible Table Builder (FTB), any inconsistencies with the published data are attributable to the fact C21CM is 98.5% of the Census 2021 whilst the FTB covers 100%. This difference is due to Census Office's coverage estimation approach which allowed for estimation and adjustment for non-response and provide a fully imputed and consistent dataset based on estimates. See the [Census 2021 Methodology](#) overview for further information.

5.0 Accessing the C21CM

5.1 Accessing the C21CM for research

Access to the C21CM dataset is limited to researchers and research teams with [accredited researcher status](#) under the Digital Economy Act 2017 (DEA) and subject to approval by the data providers. In addition, all projects must be accredited by the UK Statistics Authority Research Accreditation Panel (RAP).

In the first instance, researchers should complete the [Themed Datasets Initial Research Proposal Form](#) and submit this to rsu@nisra.gov.uk. The proposal form asks researchers to summarise the main aim and objectives of the proposed project as well as the

methodological approach that they are intending to use. Researchers must also include details of the data required (including a variable list with justification) and cohort required. Details of any public engagement activities planned as part of their project must also be included.

RSU, in consultation with the data providers, will determine the feasibility of the project prior to the completion of a [full application](#) by the researcher(s) to RAP. Ethical approval will also be required for each project, either through the research team's university ethics committee or via the [National Statistician's Ethics Advisory Committee](#). RSU will support the research team throughout the application process. Further information on how to apply for approval is available [here](#).

The C21CM dataset may be accessed in the RSU secure research environment only.

Upon approval by RAP and Census, researchers will be provided with the relevant variables in flat file tables. All research outputs will be cleared by RSU using NISRA RSU guidance

At present the C21CM is a standalone dataset and cannot be linked to other datasets.

5.2 Restricted Variables

There are a number of variables that are restricted in the C21CM due to either their sensitive nature or the level of detail, please note thorough justification will be required for these variables (examples of which are noted below). For further information about the restricted variables visit the [C21CM Metadata](#)

When completing a research proposal form to access the C21CM, please note the following due to the heightened risk of disclosure:

- if requesting any variable which is deemed sensitive (e.g. Sexual Orientation or Religion) or has a large number of categories which include sparsely populated classifications (e.g. Ethnicity full detail) then strong justification is required. This is especially the case if geographical variables are also requested.
- the more detailed variables will require a stronger justification and may in some cases require a reduced variable list. Consideration should be given to whether a pre-defined or bespoke aggregation would be more suitable.
- if multiple aggregations of the same variable are required then this too will need strong justification since combining variables in this way can lead to identification of particular classifications that are outside the standard variable coding.

- Given the low numbers involved, and the heightened risk of identifying organisations and/or establishments of sensitive nature (e.g. prisons, defence establishments) any request for multiple communal establishment (CE) variables will require strong justification. Additional to that, given the risks involved, access to CE information will require specific SDC checks to be undertaken and passed – particularly if geography variables are also selected.

6.0 Acknowledgement

When publishing or presenting findings from the C21CM the following acknowledgement should be used:

Administrative Data Research - Northern Ireland (ADR-NI) takes privacy protection very seriously. All information that directly identifies individuals/organisations will be removed from the datasets by trusted third parties before researchers access them.

All researchers using ADR-NI are trained and accredited to use sensitive data safely and ethically. They will only access the data via a secure environment and all their findings will be vetted to ensure they comply with strict confidentiality requirements.

The help provided by the staff of the Administrative Data Research Centre - Northern Ireland (ADRC-NI) and the ADR-NI support officers within NISRA Research Support Unit (RSU) is acknowledged.

ADR-NI is funded by the Economic and Social Research Council (ESRC). The authors alone are responsible for the interpretation of the data and any views or opinions presented are solely those of the author and do not necessarily represent those of ADR-NI. The data provided by the Census Office has been supplied for the sole purpose of this project.