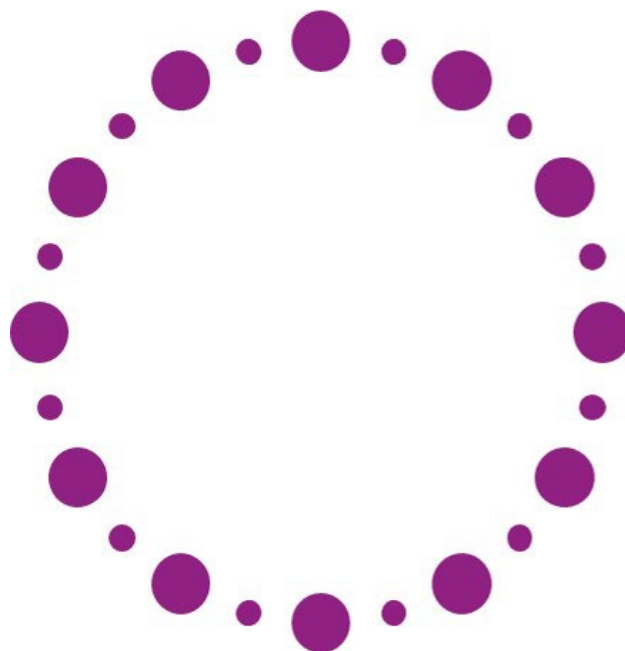


census
2021

Census 2021 Microdata strategy

21 June 2023



1. Scope

The purpose of this document is to set out the delivery plans for Census 2021 microdata products.

2. Introduction

The first results for Census 2021 statistics on the size of the Northern Ireland population were published on 24 May 2022, since then a wide range of univariate tables and bespoke multivariate tables have been published on the [Census 2021 section](#) of the NISRA website.

The Census 2021 outputs to date have made census data available for a wide range of users along with supporting information including statistical bulletins, infographics, information papers and guidance notes. A [Flexible Table Builder](#) will be made available at 9.30am on 22 June 2023, which will allow users to build Census 2021 tables for Northern Ireland on-demand. Users will be able to build custom tables by selecting the geographic level of interest and then adding individual or multiple census variables.

The next stage of Census 2021 outputs relates to microdata and is aimed at more specialist/expert users. Microdata products make it possible for researchers to access record level census data to carry out social research, and inform both the development and the monitoring of public policy. In addition, some census microdata can be linked to other data sources for approved research projects to allow specialist academic users to deliver bespoke analyses.

Due to their detailed nature, census microdata products are specifically designed to protect the confidentiality of individuals and households by using access controls, removing identifying information and applying statistical disclosure control (SDC) methods.

3. Main Census 2021 microdata products

The Census microdata products in scope for 2021 are shown in Table 1. These are products where a user need has been established, either through having been historically produced or indicated through engagement activities. Additional detail on the Census 2011 version of the products can be found in [Annex A](#).

Table 1: Census 2021 microdata products

Abbreviation	Microdata product
NILS	Northern Ireland Longitudinal Study Northern Ireland Longitudinal Study (NILS 2021)
NIMS	Northern Ireland Mortality Study 2021 Northern Ireland Mortality Study 2021 (NIMS 2021) Northern Ireland Mortality Study 2011 (NIMS 2011 extension)
SARs	Samples of Anonymised Records Teaching microdata file - 1% Sample of anonymised records (SARs 1%) Safeguarded microdata file - 5% Sample of anonymised records (SARs 5%) Secure microdata file - 10% Sample of anonymised records (SARs 10%)
OD	Origin-Destination Statistics Origin – Destination Statistics 2021
EES	Earnings and Employment Study Earnings and Employment Study 2021 (EES 2021)
CARS	Census Agriculture Research Study Census Agriculture Research Study 2021 (CARS 2021)
New – Comprehensive SAR	Comprehensive Sample of Anonymised Records [XXX%] of anonymised records (Comprehensive SAR)

4. Priorities

NISRA does not have the necessary resource to develop all microdata products simultaneously. Therefore it is necessary to prioritise their development into an appropriate order. The decision on which products to prioritise is based on a number of factors:

1. Feedback to the wide number of Census 2021 consultations (information on consultations associated with the 2021 Census is available on the [Census 2021 Consultation section](#) of the NISRA website);
2. Feedback received from direct engagement with users / researchers;
3. Consideration of the ability of the new Flexible Table Builder system to meet a significant number of users' needs;
4. Consideration of how different microdata products link together and the benefits of each; and
5. Consideration of those products that could be taken forward together (e.g. review of NILS 2021 and Comprehensive SAR variables).

Taking those factors into account, NISRA proposes the following:

- **Priority 1a:** NILS (2021) and NIMS (2021)
Aim is to launch these products in winter 2023.
- **Priority 1b:** Comprehensive SAR
Subject to approval of Northern Ireland Registrar General. Aim is to launch this product in spring 2024.
- **Priority 2:** SARs (1/5/10%) and Origin-Destination
NISRA, where possible, will align with variables and classifications included in the Office for National Statistics (ONS) releases of England and Wales microdata samples. These products would not start development until autumn 2023, and will build on lessons learnt from Priority 1 products.
- **Priority 3:** EES (2021), CARS (2021)
These products would not start development until 2024, and will build on lessons learnt from Priority 1 and Priority 2 products.

5. Other considerations

It is important to recognise that timings for the microdata products are not firm at this point as additional considerations need to be worked through. Although this may affect timing, it will increase confidence in the products delivered, including:

Geography: NISRA has developed two new statistical output geographies to support the dissemination of Census 2021 statistics. These two new statistical output geographies called Data Zones (DZ) and Super Data Zones (SDZ) are not consistent with the previous Census 2001 and Census 2011 statistical output geographies. More details on how the new geographies have been developed are available in [output geography information papers](#) on the NISRA website.

Statistical disclosure control (SDC): For Census 2021 NISRA is applying two strategies - targeted record swapping (TRS) and cell key perturbation (CKP), to ensure individuals are protected from identification in published information while minimising the impact on the quality of results. It is not yet clear how CKP can be applied to microdata products. For more information on Census 2021 SDC, the statistical disclosure control methodology is available to download on the [outputs prospectus](#) page of the NISRA website.

Comprehensive SAR concept: What does a Comprehensive SAR look like and how can it be specified in a way that works for users? For example, inclusion/exclusion of imputed items or persons, SDC swapped records, variables with known issues and different levels of variable classifications.

Future changes: There is a need to ensure consideration of any future changes and microdata developments in Office for National Statistics (ONS), National Records of Scotland (NRS) and Central Statistics Office (CSO).

New International sample: Opportunity to work with ONS to deliver an international microdata sample with the [University of Minnesota's Integrated Public-Use Microdata Series \(IPUMS\)](#) project. IPUMS is an international project that brings census microdata together from over 100 countries.

6. Main activities for Priority 1a and Priority 1b

The main activities for testing and launch of **Priority 1a - NILS 2021 and NIMS 2021** are shown in Table 2.

Table 2: Priority 1a (NILS 2021 and NIMS 2021) main activities

Activity	Description
Dataset development and creation	Define and create the NILS 2021 and NIMS 2021 datasets and metadata, including developing safe access arrangements and engaging with NILS Steering Group and researchers.
Alpha testing	Internal testing of the datasets by NISRA Research Support Unit before release to Beta testing.
Beta test project selection	Focused Beta test project development, application and assessment. Beta test project teams will include experienced NILS/NIMS researchers drawn from the user community and within NISRA. Project applications will be assessed by NISRA Census Office and Research Support Unit against agreed key test criteria in August 2023.
Beta test data preparation	Dataset builds for Beta test projects and initiation meetings with all Beta test teams.
Beta testing	Live running of Beta test projects, with access to Beta test database in NISRA Research Support Unit Secure Room. It is planned for Beta projects to run in October and November 2023.
Evaluation and launch decision	Review of Beta testing, any refinements based on testing and formal decision on whether to proceed with launch. Researchers able to publish Beta test outputs with launch.

After Launch the proposed access to NILS 2021 and NIMS 2021 project databases will be the same as that for Beta testing and NILS 2011 - researchers would have access to project databases in the NISRA Research Support Unit Secure Room. The Census 2021 Flexible Table Builder system will assist users as it will allow users to explore Census 2021 data outside the safe room.

The main activities for testing and launch of **Priority 1b - Comprehensive SAR** are shown in Table 3.

Table 3: Priority 1b (Comprehensive SAR) main activities

Activity	Description
Development of concept and processes	Engage with ONS and NRS, working towards harmonising the approach to product content and access arrangements/policies as much as possible.
Dataset development and creation	Define and create the Comprehensive SAR dataset and metadata, including developing access arrangements, project approvals processes and engaging with potential users.
Test project selection	Test project processes still to be developed, opportunity to use lessons learnt from NILS/NIMS test project selection.
Testing	Live running of test projects, processes still to be developed, and opportunity to use lessons learnt from NILS/NIMS testing. With access to test databases in NISRA Research Support Unit Secure Room.
Evaluation and launch decision	Review of testing, any refinements based on testing and formal decision on launch.

Proposed access for testing and initial projects after launch through the NISRA Research Support Unit Secure Room.

Annex A – Additional census microdata detail

More details on the 2011 Census microdata products are shown in Table 4.

Table 4: Additional detail on the 2011 Census microdata products

Product	Description	Access	Proposal for 2021
Northern Ireland Longitudinal Study (NILS 2011)	The Longitudinal Study (NILS) is a dataset incorporating a sample of data from the 1981, 1991, 2001 and 2011 Censuses. Further information is available on the Northern Ireland Longitudinal Study research support pages of the NISRA website.	Secure - Restricted to NISRA safe room.	Census Office will enhance this resource with data from Census 2021.
Northern Ireland Mortality Study (NIMS 2011)	The NIMS 2011 contains links for the enumerated population of Northern Ireland from the 2011 Census to registered mortality data from the General Register Office (GRO). Further information is available on the Northern Ireland Longitudinal Study research support pages of the NISRA website.	Secure - Restricted to NISRA safe room.	No significant changes planned – only minor changes to variables based on changes made to the 2021 Census questionnaire.
Census 2011 - Teaching microdata file (1% SARs)	A publicly available microdata file containing anonymised records on a limited set of variables for a random sample of 1% of individuals from Census 2021 for Northern Ireland. The 2011 teaching file is available on the microdata teaching file page of the NISRA website. 1 file – Individuals, region level, approx. 20 variables	Public - Available to all on the NISRA website.	No significant changes planned – only minor changes to variables based on changes made to the 2021 Census questionnaire.

Product	Description	Access	Proposal for 2021
<p>Census 2011 - Safeguarded individual microdata file (5% SARs)</p>	<p>A safeguarded microdata file which will consist of a random sample of 5% of individuals from Census 2021 for Northern Ireland. More information on the 2011 5% SARs is available at safeguarded microdata page of the NISRA website.</p> <p>2 files - 1st Individuals, region level, approx. 120 variables, 2nd Individuals, LGD level, approx. 120 variables</p>	<p>Safeguarded - Available to data analysts under certain terms and conditions, 2011 version hosted on the UK Data Service.</p>	<p>No significant changes planned – only minor changes to variables based on changes made to the 2021 Census questionnaire.</p>
<p>Census 2011 - Secure microdata file (10% SARs)</p>	<p>The secure microdata file will consist of two random samples of 10%, one a sample of the individuals and one a sample of households from the 2021 Census for Northern Ireland. More information on the 2011 10% SARs are available secure microdata page of the NISRA website.</p> <p>2 files - 1st Individuals, LGD level, approx. 200 variables, 2nd HHs, LGD level, approx. 200 variables</p>	<p>Secure - Only available to approved or accredited researchers, 2011 version hosted on the Office for National Statistics Secure Researcher Service and in the NISRA safe room.</p>	<p>No significant changes planned – only minor changes to variables based on changes made to the 2021 Census questionnaire.</p>

Product	Description	Access	Proposal for 2021
Origin – Destination Statistics 2011	Origin-destination data from the census provides aggregate statistics from census returns on the movement of people from one location to another. Examples include migration (address one year ago) and workplace or place of study flows. More information can be found on the origin-destination pages on the Office for National Statistics website.	Range of datasets with three different access types, from Public (available to all), to Safeguarded (available to data analysts under certain terms and conditions on the UK Data Service), to Secure (only available to approved or accredited researchers on the Office for National Statistics Secure Researcher Service or in the NISRA safe room).	No significant changes planned – only minor changes to variables based on changes made to the 2021 Census questionnaire.
Earnings and Employees Study (EES 2011)	The EES links together variables from the Annual Survey of Hours and Earnings with variables from the Census, and Capital Value data from the Land and Property Services. For use in research to inform development and monitoring of public policy and provide an evidence base for decision making, using factors influencing variation in median pay, graduates and the workplace, disparity in pay and migration, public versus private jobs, and geography and pay. More information on the 2011 version is available on its themed dataset page on the NISRA website.	Secure - Only available to approved or accredited researchers on the Office for National Statistics Secure Researcher Service or in the NISRA safe room.	Consideration will be given to an enhanced variable set for the 2021 EES and also for a Longitudinal Earnings element – although this will be not in the initial phase.

Product	Description	Access	Proposal for 2021
<p>Census Agriculture Research Study 2021 (CARS 2011)</p>	<p>Census Agriculture Research Study 2021 (CARS 2021), would be an update for the CARS 2011 dataset.</p> <p>CARS 2011 links de-identified Census 2011 records with agricultural administrative data. It is part of a larger UK project focusing on the composition and characteristics of farm households and undertaking analyses investigating farm household health and wellbeing, prosperity and resilience, and farmer engagement with agri-environmental issues. More detail on the CARS (2011) can be found on the ADR UK project webpage.</p>	<p>Secure - Restricted to NISRA safe room.</p>	<p>No significant changes planned.</p>