

A unified dataset of UK census variables for small areas: Harmonised data tables from the 2021 England, Wales, and Northern Ireland censuses and the 2022 Scotland census

EPB: Urban Analytics and City Science
2026, Vol. 0(0) 1–13
© The Author(s) 2026



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/23998083261429563

journals.sagepub.com/home/epb



Owen Goodwin¹  and Alex Singleton¹ 

Abstract

This paper describes a new dataset release containing harmonised census tables from the 2021 and 2022 UK Censuses. The release is the first unified dataset covering all four UK nations at the smallest available geographic level: Output Areas in England, Wales, and Scotland, and Data Zones in Northern Ireland. The UK's three census agencies: ONS (England and Wales), NRS (Scotland), and NISRA (Northern Ireland) release their data separately, each with distinct variables, formats, and disclosure controls. Through a process of matching, standardisation, and aggregation, 190 comparable variables are produced. The dataset is made available as a series of topic tables indexed across all 239,023 of the UK's small-area geographies. By providing a standardised dataset, this work enables seamless UK-wide analyses, facilitating cross-national comparisons and supporting research and public policy development.

Keywords

census, census 2021, United Kingdom

Introduction

Census data remain one of the most important sources of demographic information for public policy decision making and academic research (Killick et al., 2016) as they generate insight that captures

¹Department of Geography and Planning, University of Liverpool, Liverpool, UK

Corresponding author:

Owen Goodwin, Department of Geography and Planning, University of Liverpool, Digital Innovation Facility, Liverpool L69 3RF, UK.

Email: owen.goodwin@liverpool.ac.uk

Data Availability Statement included at the end of the article

the demographic, social, and economic characteristics of individuals, households, and neighbourhoods. In addition to underpinning academic research and policy making in the UK, given their open license, the availability of these data encourages the creation of derived data products such as nationwide socio-economic classifications or indicators (Stillwell, 2017; Wyszomierski et al., 2024).

The conduct and administration of the 2021/22 population censuses across the UK varied between its constituent countries. The four countries of the UK were covered by three separate censuses, conducted by the Office of National Statistics for England and Wales, the National Records of Scotland (NRS) for Scotland, and the Northern Ireland Statistics and Research Agency (NISRA) for Northern Ireland. Census data are released separately by the respective agencies after a period of processing and disclosure control.

Working with census data on a UK-wide basis has been a long-standing challenge across multiple census rounds. Differences across the constituent nations in how census data are collected, processed, and released create barriers to research requiring unified UK-wide data, necessitating expert knowledge to navigate these complexities. This is illustrated by work harmonising census migration and commuting interaction data (Stillwell and Duke-Williams, 2007) and producing consistent estimates of cross-border migration within the UK (Lomax et al., 2013). For the 2011 census, a unified dataset was produced by the UK Data Service to enable researchers to work more readily with UK-wide census data (Dymond-Green, 2017). However, to date, no equivalent unified dataset has been produced for the 2021/22 census.

The 2021 Census of England, Wales, and Northern Ireland occurred on 21 March 2021, and the Census of Scotland took place on 20 March 2022. The delay from 2021 to 2022 in Scotland was a result of local decisions to mitigate the effects of the COVID-19 pandemic. However, despite the extra year of preparation, Scotland’s Census initially faced lower-than-expected response rates than in other parts of the UK conducted a year earlier. This required an extension to the survey window and additional targeting (Table 1). Despite such intervention, the overall response rate was still lower than anticipated. As such, the creation of estimates from the census involved implementing more extensive bias-mitigation measures than for the rest of the UK (National Records of Scotland, 2024). Furthermore, the separate timings in Scotland have been noted to have introduced challenges for public awareness and engagement (Audit Scotland, 2022).

The Office for Statistics Regulation noted that conducting a census at the height of COVID-19 pandemic in 2021 has meant that some of the topic data would reflect the pandemic conditions or controls. In particular, statistics on employment, economic activity, commutes, and household structures “may well be unusual or changed” in 2021 (Office for Statistics Regulation, 2025a; 2025b). Similar issues were also noted in Scotland despite the 12-month delay (National Records of Scotland, 2024). Without extensive resurveying, it is not possible to control for such issues endogenously. As such, our approach for the unified UK-wide census dataset is to acknowledge these issues within the underlying survey and highlight them in the

Table 1. Census response and online completion rates (National Records of Scotland, 2024; Northern Ireland Statistics and Research Agency, 2022; Office for National Statistics, 2022).

Country	Response rate (%)	Online rate (%)
England and Wales	97	88
Northern Ireland	97	80
Scotland	79 (initial); 90 (final)	90

metadata of tables where they are most acute. There is also the general caveat outside of pandemic-specific issues in that there is a date misalignment between 2021 and 2022.

Within this context, our paper presents reproducible code that has been used to create an integrated UK-wide 2021/22 small-area census dataset. Each table that we have created includes detailed notes describing the compatibility of the variables across the three censuses, outlining cases where aggregations or adjustments were necessary due to differences in definitions, classifications, or response categories. The resulting data product facilitates UK-wide research while maintaining a clear link to the original census data sources.

2021/2022 UK census compatibility and comparability

The release of small-area census tables occurs separately for each UK nation, following the independent publication schedules of their respective statistical agencies. Outside of the issues related to the Pandemic, many of the separate small-area census tables released are broadly comparable between countries. However, some important differences emerge in terms of the questions asked, variable descriptions provided, disclosure controls, and release formats. Such factors create obstacles for less expert users of census data in UK-wide analyses, and acutely so, given that many key differences are not immediately apparent. As an illustrative example, the ONS and NRS did not include a question about the number of rooms in a household, asking only about the number of bedrooms due to questionnaire space constraints. NISRA did not ask about either rooms or bedrooms and instead used Land & Property Services data to construct a variable for the total number of rooms. In England and Wales, the Valuation Office Agency provided an alternative source of data for the number of rooms. Scotland, however, sourced their response from the survey and did not use supplementary data to construct a total room variable. As a result, there are no fully equivalent measures for the number of rooms or number of bedrooms across the UK, which has been available in previous censuses ([Stillwell, 2017](#)).

Furthermore, variable categorisations differ between countries for certain questions, driven by differences in demographic makeup that impact disclosure controls and by the priorities of the respective nations. For example, published small-area tables related to ethnicity have significant discrepancies between the nations. For instance, both England and Scotland include a distinct category for individuals identifying as having Bangladeshi ethnicity, whereas in Northern Ireland, this category is absent due to the low proportion of this population there. Conversely, Northern Ireland includes a category for individuals of Filipino ethnicity, which is absent from the classifications used in England, Wales, and Scotland, reflecting the higher proportion of this group within Northern Ireland. Additionally, the categorisation of Irish Travellers differs between the nations: Northern Ireland records “Irish Traveller” as a distinct ethnic category from “White.” In contrast, in England, Wales, and Scotland, the variable “Irish Traveller” is nested within the broader “White” category. While such differences are logical within the context of the different nations, they are an additional challenge for the analyst wishing to harmonise variables for UK-wide analysis.

In practical terms, a researcher who is currently interested in UK-wide analysis using the 2021 and 2022 small-area census data must access each of the three census data extracts separately and verify that the tables of interest contain comparable variables. This may often require a direct comparison of census survey questions. To add further complexity, table and variable naming and formatting can be inconsistent even when derived from identical census questions. This can be particularly problematic for applications that require data to be machine-readable.

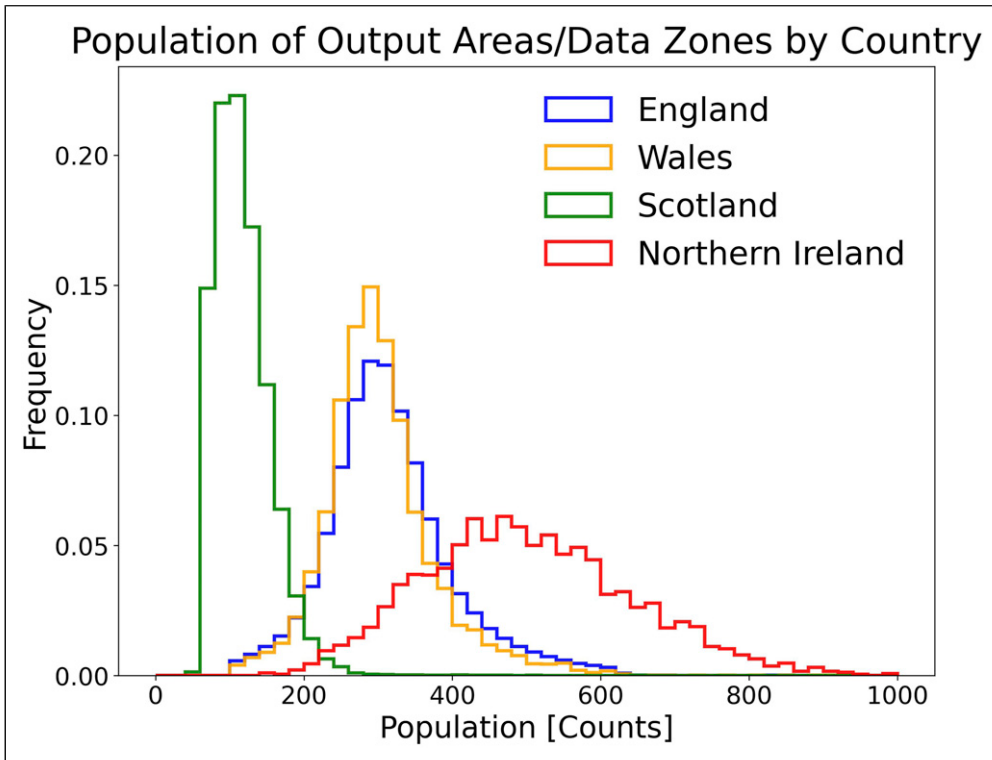


Figure 1. Distributions of the population size of output areas/Data zones for each of the countries of the United Kingdom.

This paper therefore presents a harmonised UK-wide set of census variables at the smallest set of available geographies: Output Areas (called Data Zones in Northern Ireland). These areas, first implemented following the 2001 Census, serve as the building blocks for other census geographies. They are designed to be demographically homogeneous units, while also respecting existing natural and administrative boundaries as much as possible (Martin, 2002). Within each nation, Output Areas are designed to have roughly equal populations, but the population target size varies between nations due to differences in population distribution and administrative needs. The population size distribution, by country, can be seen in Figure 1. In England and Wales, Output Areas contain around 125 households (≈ 300 people), while in Scotland, they are smaller, averaging 50 households (≈ 125 people). In Northern Ireland, the equivalent units, Data Zones, are larger, with around 200 households (≈ 500 people).

Dataset construction and harmonisation methods

Input data were derived from the most recent Census release for each of the four UK nations, obtained from open-access sources. Data for England and Wales were accessed via the ONS' Nomis bulk download portal.¹ Scottish census data were obtained as a bulk table download from Scotland's Census website,² as provided by the NRS. For Northern Ireland, data were extracted from the NISRA Flexible Table Builder³ through automated queries. As no API was available, automated queries were performed using web scraping to retrieve the available tables, allowing

Table 2. The number of available topic tables for each of the individual census releases and in the harmonised dataset.

	Number of tables
England and Wales tables	52
Scotland tables	62
Northern Ireland tables	279
Table Match candidates	28
— With incompatible variables	— 3
— With compatible variables	— 25
— With <i>directly matching variables</i>	— 4
— With <i>variables requiring aggregation</i>	— 21

direct access to the download URLs for each table. The analysis is restricted to univariate tables, due to limitations in data availability; at the Output Area level, multivariate data are not consistently available, as small population counts often lead to suppression or disclosure control measures.

Variable harmonisation

Minor data inconsistencies were first addressed by removing repeated variables and correcting some typographical errors in the variable descriptions. All datasets were standardised into a common tabular structure, with counts for each variable labelled by a unique code and indexed by Output Area/Data Zone code (ONS Geography, 2009). Each variable ID was formatted *AA**XXX**PPPP*, where *AA* is the country code of the dataset, *XXX* is the table ID, and *PPPP* is the sequential position of the variable in the table. Table totals are the first variable in each table (with ID *AA**XXX**0001*). The country codes are *ts* for England and Wales, *uv* for Scotland, and *ni* for Northern Ireland, with *ts* and *uv* chosen to match the original table ID formats in the respective datasets. Table IDs were selected to correspond to those in the original releases for England, Wales, and Scotland, ensuring consistency with other resources. As table codes do not exist in the Northern Ireland release, they were created sequentially to maintain a structured format. Consistent metadata tables were generated as lookup tables for variable codes, providing full variable descriptions, units, and source information. Variables are organised into topic tables matching the structure of the original data releases, resulting in 52 topic tables for England and Wales, 62 for Scotland, and 279 for Northern Ireland. There were significantly more tables for Northern Ireland as NISRA does not use sub-variables, instead releasing separate tables with different levels of variable aggregation for each topic table.

A manual process was undertaken to identify comparable tables across the three census releases. When multiple Northern Ireland tables existed for the same topic with different levels of categorisation, the variables were reviewed to select the table with the most detailed categorisation which aligned with the variables used in the other countries. The outcome of the table matching is shown in Table 2. As an indication of the scale of the difference, of 52 available tables in the ONS output area release, only 28 tables are matched across the three census releases. Of these matched tables, three were deemed to have incompatible variables.

Of the 25 tables with matched candidate variables, only four comprised entirely comparable variable categorisations. For the other tables where different levels of categorisation were used, variables were aggregated to create directly comparable sets of variables, with the

Table 3. Schematic showing the variable assignment and aggregation process for the unified “Ethnic Group” table (uk021) using variables from England and Wales (E & W, table ts021), Scotland (SC, table UV201), and Northern Ireland (NI, table ni060).

UK Unified (uk021)	E & W (ts021)	SC (UV201)	NI (ni060)
0001: Total: All usual residents	0001: Total: All usual residents	0001: All people	0001: Table total
0002: Asian	0002: Asian, Asian British or Asian Welsh	0010: Asian, Asian Scottish or Asian British: Total	0005: Indian + 0006: Chinese + 0007: Filipino + 0008: Pakistani + 0010: Other Asian
0003: ...: Bangladeshi	0003: ...: Bangladeshi	0013: ...:Bangladeshi, Bangladeshi Scottish or Bangladeshi British	—
0004: ...: Chinese	0004: ...: Chinese	0014: ...: Chinese, Chinese Scottish or Chinese British	0006: Chinese
0005: ...: Indian	0005: ...: Indian	0012: ...: Indian, Indian Scottish or Indian British	0005: Indian
0006: ...: Pakistani	0006: ...: Pakistani	0011: ...: Pakistani, Pakistani Scottish or Pakistani British	0008: Pakistani
0007: ...: Other Asian	0007: ...: Other Asian	0015: ...:Other Asian	0007: Filipino + 0010: Other Asian
0008: Black, Caribbean or African	0008: Black, Black British, Black Welsh, Caribbean or African	0016: African: Total + 0019: Caribbean or Black: Total	0011: Black African + 0012: Black Other
0009: Mixed or multiple ethnic groups	0012: Mixed or Multiple ethnic groups	0009: Mixed or multiple ethnic groups	0013: Mixed
0010: White	0017: White	0002: White: Total	0002: White + 0003: Irish traveller + 0004: Roma
0011: Other ethnic group	0023: Other ethnic group	0023: Other ethnic groups: Total	0009: Arab + 0014: Other ethnicities
0012: ...: Arab	0024 ...: Arab	0024: ...: Arab, Arab Scottish or Arab British	0009: Arab
0013: ...: Any other ethnic group	0025: ...: Any other ethnic group	0025: ...: Other ethnic group	0014: Other ethnicities

Only variables that are matched or aggregated to unified variables are shown. The “+” symbol indicates aggregation of multiple source categories and “...” denotes subcategories within broader ethnic groups.

objective for these to be as detailed as possible. However, this process still inevitably reduced granularity. [Table 3](#) shows an example schematic indicating the variable matching and aggregation performed for the ethnic group table discussed in the previous section. Aggregations were performed such as combining “Ethnic Group: Filipino,” which only appears in the Northern Ireland data, into the “Ethnic Group: Other Asian” variable. Additionally, differences in the subcategories between Scotland and Northern Ireland meant that Scotland’s distinct “Caribbean or Black” and “African” totals and Northern Ireland’s “Black African” and “Black Other” variables were merged into a broader “Black, Caribbean or African” variable. Due to the differences in question structure, Northern Ireland’s separate “Irish Traveller” and “Roma” variables were combined with “White” to match the aggregation in other nations. Full details of all the aggregations performed and their explanations are included in the data release.

Statistical disclosure control harmonisation

To ensure the anonymity of individuals and households in the censuses, the agencies each carry out a process of statistical disclosure control. The released data is anonymised using two key methods: “Targeted record swapping” and “Cell key perturbation.” Targeted record swapping involves a small percentage of households having their census records swapped with similar households in different geographic areas, with households containing individuals with rare or unique characteristics who might be easier to identify being more likely to be swapped. Cell key perturbation involves small random adjustments being applied to cell counts to prevent the identification of individuals, with smaller cells being more likely to receive larger perturbations ([National Records of Scotland, 2020a, 2020b](#); [Northern Ireland Statistics and Research Agency, 2021](#); [Office for National Statistics, 2023](#)).

The input tables for England, Wales, and Scotland used in this dataset have passed additional checks to identify sparse tables which could still be disclosive. The flexible table builder we have used to extract outputs for Northern Ireland suppresses sparse tables at the Data Zone level. This does not apply to any of the univariate tables we have extracted. Therefore, small cell count suppression has not been applied to any of the input variables from any of the agencies. Instead, all cell counts are protected through the perturbation methods described above, meaning that every cell contains a value (including zeros), albeit with small random adjustments applied for disclosure control purposes. The unified tables therefore contain all OAs within each nation, ensuring complete geographic coverage for the released dataset.

While each of the agencies follows a similar disclosure control process, there are differences in how the data are tabulated afterwards. After performing cell key perturbation, the ONS sums the new perturbed values to calculate a new total for that table, whereas the NRS provides the original totals. This leads to occurrences in the Scottish data where for some smaller Output Areas the table total is less than the sum of the sub-variables. Northern Ireland does not include totals in the tables produced by the Flexible Table Builder from which we extract data. To ensure consistency across the unified dataset, we follow the approach used by the ONS for England and Wales: we calculated totals for Northern Ireland based on the sub-variables provided and recalculated Scottish totals, replacing the original values with the sum of the perturbed sub-variables. This totalling procedure is also applied in cases where a table variable is a subtotal, that is, the sum of two or more other table variables.

Table 4. The topic tables included in the harmonised UK release.

UK table ID	Table name	Unit	Number of variables
uk001	Residence type	Person	2
uk002	Marital and civil partnership status	Person	5
uk003	Household composition	Household	13
uk004	Country of birth	Person	8
uk007	Age	Person	18
uk008	Sex	Person	2
uk015	Year of arrival in the UK	Person	12
uk017	Household size	Household	5
uk021	Ethnic group	Person	12
uk023	Multiple ethnic groups in household	Household	5
uk029	Proficiency in English language	Person	2
uk030	Religion	Person	6
uk037	General health	Person	5
uk038	Disability	Person	4
uk039	Provision of unpaid care	Person	5
uk044	Accommodation type	Household	7
uk045	Number of cars or vans	Household	4
uk046	Type of central heating in household	Household	10
uk054	Tenure of household	Household	8
uk059	Hours worked	Person	6
uk061	Method of travel to workplace	Person	10
uk062	National statistics socio-economic classification (NS-SEC)	Person	9
uk063	Occupation (current)	Person	9
uk066	Economic activity status	Person	18
uk067	Highest level of qualification	Person	5

Due to this approach, the table totals reflect the sum of the perturbed variables. Therefore, totals of households or population will differ across tables even when they conceptually refer to the same measure. In particular, the totals from individual tables should not be interpreted as definitive “single number” estimates of the population or number of households in an OA. This characteristic has been a long-standing feature of ONS census releases ([Rees et al., 2005](#)).

Data release

In total, 190 unified variables were produced and organised into 25 thematic tables. The tables are provided in both CSV and parquet formats.⁴ The available tables are listed in [Table 4](#). Tables either contain counts of individual populations (Person) or entire households (Household). The tables are indexed by the Output Area (OA) or Data Zone (DZ) code, using a total of 239,023 unique small-area codes. A note file is provided which describes any notable features in the harmonisation of each table. Variable IDs are created following the same code structure as the standardised input tables. Here, the unified country code is designated as “UK,” and the table IDs are the corresponding IDs from the England and Wales release.

As an additional output, the Python code used to produce the dataset is made available for download.⁵ This includes all stages of dataset construction, including accessing the original separate census data, processing these into consistent formats, and combining variables across censuses to produce the final

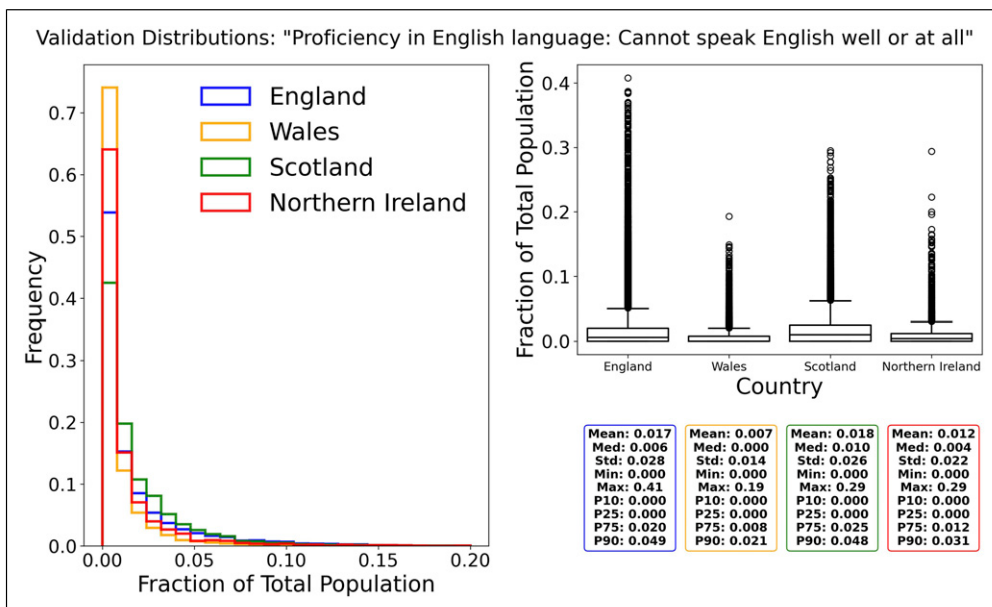


Figure 2. Validation distributions of the aggregated variable “proficiency in English language: Cannot speak English well or at all.” Left: The frequency distribution of the fraction of the total small-area population for this variable, with separate histograms for each country. Right: Boxplots summarising the distributions with key statistics: the mean, median, standard deviation, minimum, maximum, and percentiles (10th, 25th, 75th, and 90th) displayed.

outputs. The code provides a fully reproducible workflow for recreating the unified data, as well as documented lookup tables between variables in the unified data and the original separate census releases.

Internal validation

The harmonisation process was checked by examining summary statistics and the distributions of the unified variables by country. The summary statistics included the mean, median, quartiles, and range. The validation plots for all variables are available in the code repository. The overarching purpose of this validation process was to examine any differences between countries that might have been a result of erroneous assumptions or mistakes in the matching process.

Figure 2 presents an example validation distribution for the variable: “Proficiency in English language: Cannot speak English well or at all.” This variable is an aggregate measure derived from different census questions across the UK nations. In England and Northern Ireland, the Census included a question specifically about spoken English ability, with response options: *Very well*, *Well*, *Not well*, and *Not at all*. In contrast, Scotland’s Census asked separate questions about individuals’ ability to understand, read, write, and speak English. As a result, Scotland released a nine-category classification of overall English proficiency, incorporating elements from all four skills. Based on the descriptions of these variables, it was determined that two compatible measures could be derived through aggregation: “Can speak English very well or well” and its inverse, “Cannot speak English

Table 5. Schematic showing the variable assignment and aggregation process for the unified proficiency in English language table (uk029) using variables from England and Wales (E&W, table ts029), Scotland (SC, table UV210), and Northern Ireland (NI, table ni056).

UK (uk029)	E&W (ts029)	SC (UV210)	NI (ni056)
0001: Total: All usual residents aged 3 years and over	0001: Total: All usual residents aged 3 years and over	0001: All people aged 3 and over	0001: Table total
0002: Speaks English very well or well	0002: Main language is English (English or Welsh in Wales) +	0003: Speaks, reads and writes English +	0002: Main language is English +
	0004: Main language is not English: Can speak English very well +	0004: Speaks but does not read or write English +	0003: Main language is not English: can speak English very well +
	0005: Main language is not English: Can speak English well	0005: Speaks and reads but does not write English	0004: Main Language is not English: Can speak English well
0003: Cannot speak English well or at all	0006: Main language is not English: Cannot speak English well +	0002: Understands spoken English only +	0005: Main language is not English: cannot speak English well +
	0007: Main language is not English: Cannot speak English	0006: Reads but does not speak or write English +	0006: Main language is not English: cannot speak English
		0007: Writes but does not speak or read English +	
		0008: Reads and writes but does not speak English +	
		0010: Limited English skills +	
		0011: No skills in English	

Only variables that are matched or aggregated to unified variables are shown. The “+” symbol indicates aggregation of multiple source categories.

well or at all.” Table 5 presents a schematic documenting the aggregation used to produce these unified variables.

The validation plots demonstrate that these aggregated variables are broadly compatible in their distributions across the different census datasets. As discussed earlier, for such complex cases, details and considerations are included in our metadata so that these can be taken into account when judging the suitability of a table for a particular type of analysis.

Conclusion

The harmonised dataset described in this paper represents the first unified source of 2021/22 small-area census data across the UK, addressing key challenges in data comparability between the Censuses of England, Wales, Scotland, and Northern Ireland. The harmonisation process exposes differences in variable categorisation, disclosure controls, and the use of external data across the four nations. These differences underscore the challenges of working with census data on a cross-national basis.

While this dataset is UK-specific, harmonisation of census and similar population data is a challenge across a wide range of international contexts. Similar difficulties arise in transnational policy evaluation, such as efforts to harmonise and centralise census data across EU member states (Bach, 2019; Pertiwi and Nugrahani, 2020), and in research throughout federated countries where data from multiple jurisdictions must be linked, for example, in Australia (Boyd et al., 2012; Rosman et al., 2016) and Canada (Katz et al., 2018). The fully open and reproducible methodological framework presented here, including the code pipeline with systematic variable mapping, transparent documentation of aggregation decisions, and validation procedures, could be adapted to other multi-jurisdictional harmonisation contexts.

There are constraints on the degree of harmonisation that can be achieved using the existing openly available data. Differences in release table structures between nations limit the number of variables available in the harmonised dataset. Even when the underlying census question is identical, country-specific decisions on how tables are categorised lead to cases where significant aggregation of individual variables is necessary to achieve a comparable set of variables. An additional limitation arises from the separate application of the statistical disclosure control methods of cell key perturbation and target record swapping, which are performed at the country level before harmonisation. When aggregation is performed, combining multiple values that have been individually perturbed or swapped introduces further distortion. These constraints would not apply if harmonisation was performed on the original data. UK statistical agencies could collaborate to establish an official standardised set of harmonised census outputs. However, in the absence of such a release, the dataset presented here remains the most accessible and comprehensive solution for cross-national 2021/22 census analysis.

By providing 190 harmonised variables across 239,023 small-area geographies, this dataset enables UK-wide demographic analysis at the most granular level that is openly available. Each variable has been individually validated to ensure compatibility. The inclusion of detailed metadata and an open-access code release facilitate reproducibility and accessibility, ensuring that this dataset can be readily utilised for research and policy development.

ORCID iDs

Owen Goodwin  <https://orcid.org/0000-0002-2592-1338>

Alex Singleton  <https://orcid.org/0000-0002-2338-2334>

Ethical considerations

No ethical approval required.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Economic and Social Research Council [ES/Z504464/1, ES/Z50273X/1]

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

The UK unified census data product is available to download from the Geographic Data Service at <https://data.geods.ac.uk/dataset/unified-uk-census-data>. The code pipeline to produce the product is available at <https://github.com/GeographicDataService/unified-uk-census-2021-22/>.

Notes

1. https://www.nomisweb.co.uk/sources/census_2021_bulk.
2. <https://www.scotlandscensus.gov.uk/documents/2022-output-area-data/>.
3. <https://www.nisra.gov.uk/statistics/results/flexible-table-builder>.
4. <https://data.geods.ac.uk/dataset/unified-uk-census-data>.
5. <https://github.com/GeographicDataService/unified-uk-census-2021-22/>.

References

- Audit Scotland (2022) *The 2021/22 Audit of National Records of Scotland* | Audit Scotland. Available at: <https://audit.scot/publications/the-202122-audit-of-national-records-of-scotland> (Accessed 25 March 2025).
- Bach F (2019) Statistical disclosure control in geospatial data: the 2021 EU census example: 365–384. Available at: https://doi.org/10.1007/978-3-319-72434-8_18
- Boyd JH, Ferrante AM, O’Keefe CM, et al. (2012) Data linkage infrastructure for cross-jurisdictional health-related research in Australia. *BMC Health Services Research* 12(1): 480, Available at: <https://doi.org/10.1186/1472-6963-12-480>
- Dymond-Green R (2017) *Creating a Unified 2011 Census Dataset for the Four Nations of the UK*. Available at: <https://blog.ukdataservice.ac.uk/creating-a-unified-2011-census-dataset-for-the-four-nations-of-the-uk/> (Accessed 23 October 2025).
- Katz A, Enns J, Wong ST, et al. (2018) Challenges associated with cross-jurisdictional analyses using administrative health data and primary care electronic medical records in Canada. *International Journal of Population Data Science* 3(3): 437, Available at: <https://doi.org/10.23889/ijpds.v3i3.437>
- Killick L, Hall H, Duff AS, et al. (2016) The census as an information source in public policy-making. *Journal of Information Science* 42(3): 386–395, Available at: <https://doi.org/10.1177/0165551516628471>
- Lomax N, Norman P, Rees P, et al. (2013) Subnational migration in the United Kingdom: producing a consistent time series using a combination of available data and estimates. *Journal of Population Research* 30(3): 265–288, Available at: <https://doi.org/10.1007/s12546-013-9115-z>
- Martin D (2002) Geography for the 2001 census in England and Wales. *Population Trends* 108: 7–15.
- National Records of Scotland (2020a) *Scotland’s Census 2022 Cell Key Perturbation*. Available at: <https://www.scotlandscensus.gov.uk/media/d1yn5gu3/pmp017-cell-key-perturbation-emap-5940.pdf> (Accessed 20 March 2025).
- National Records of Scotland (2020b) *Scotland’s Census 2022 Statistical Disclosure Control & Outputs*. Available at: <https://www.scotlandscensus.gov.uk/media/dqbn3hfr/pmp016-household-record-swapping-emap-5938.pdf> (Accessed 20 March 2025).
- National Records of Scotland (2024) *Scotland’s Census 2022 - General Report*. Available at: <https://www.scotlandscensus.gov.uk/about/scotlandscensus-2022-general-report/> (Accessed 24 March 2025).
- Northern Ireland Statistics and Research Agency (2021) *Statistical Disclosure Control Methodology for 2021 Census*. Available at: <https://www.nisra.gov.uk/files/nisra/publications/statistical-disclosure-control-methodology-for-2021-census.pdf> (Accessed 11 April 2025).
- Northern Ireland Statistics and Research Agency (2022). *Census 2021 Statement About Data Quality*. Available at: <https://www.nisra.gov.uk/publications/census-2021-statement-about-data-quality> (Accessed 11 April 2025).
- Office for National Statistics (2022) *Maximising the Quality of Census 2021 Population Estimates*. Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/maximisingthequalityofcensus2021populationestimates> (Accessed 25 March 2025).

- Office for National Statistics (2023) *Protecting Personal Data in Census 2021 Results*. Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/protectingpersonaldataincensus2021results> (Accessed 20 March 2025).
- Office for Statistics Regulation (2025a) *Assessment of Compliance With the Code of Practice for Statistics – 2021 Census in England and Wales*, Office for Statistics Regulation. Available at: <https://osr.statisticsauthority.gov.uk/publication/assessment-of-compliance-with-the-code-of-practice-for-statistics-2021-census-in-england-and-wales/> (Accessed 24 March 2025).
- Office for Statistics Regulation (2025b) *Assessment of Compliance With the Code of Practice for Statistics – 2021 Census in Northern Ireland*, Office for Statistics Regulation. Available at: <https://osr.statisticsauthority.gov.uk/publication/assessment-of-compliance-with-the-code-of-practice-for-statistics-2021-census-in-northern-ireland/> (Accessed 24 March 2025).
- ONS Geography (2009) *Coding and Naming Policy for UK Statistical Geographies*. Available at: <https://geoportal.statistics.gov.uk/documents/coding-and-naming-policy-for-uk-statistical-geographies-1/about> (Accessed 26 March 2025).
- Pertiwi E and Nugrahani HD (2020) Census hub as centralizing population data: the European union experience. *Journal of Strategic and Global Studies* 3(2): Article 4, Available at: <https://doi.org/10.7454/jsgs.v3i2.1033>
- Rees P, Parsons J and Norman P (2005) Making an estimate of the number of people and households for output areas in the 2001 census. *Population Trends* 122: 27–34.
- Rosman D, Spilsbury K, Alan J, et al. (2016) Multi-jurisdictional linkage in Australia: proving a concept. *Australian and New Zealand Journal of Public Health* 40(1): 96–97, Available at: <https://doi.org/10.1111/1753-6405.12420>
- Stillwell J (ed) (2017) *The Routledge Handbook of Census Resources, Methods and Applications: Unlocking the UK 2011 Census*. 1st edition. Routledge. Available at: <https://doi.org/10.4324/9781315564777>
- Stillwell J and Duke-Williams O (2007) Understanding the 2001 UK census migration and commuting data: the effect of small cell adjustment and problems of comparison with 1991. *Journal of the Royal Statistical Society Series A: Statistics in Society* 170(2): 425–445, Available at: <https://doi.org/10.1111/j.1467-985X.2006.00458.x>
- Wyszomierski J, Longley PA, Singleton AD, et al. (2024) A neighbourhood output area classification from the 2021 and 2022 UK censuses. *The Geographical Journal* 190(2): e12550, Available at: <https://doi.org/10.1111/geoj.12550>

Author biographies

Owen Goodwin is a Senior Data Scientist at the University of Liverpool. His research focuses on data science methodologies applied to large-scale geospatial datasets.

Alex Singleton a Professor of Geographic Information Science at the University of Liverpool. His research is concerned with how the complexities of individual behaviours, attitudes, and contexts manifest spatially and can be represented and understood through a framework of Geographic Data Science.