Understanding Society: Geographical Data

USER GUIDE

Version 1, September 2025

Understanding Society

Economic and Social Research Council

Contents

1. Introduction to Understanding Society Geographical datasets

In this guidance document, the geographical identifiers and geographical classifications datasets available in the main Understanding Society study (UK Household Longitudinal Study, UKHLS), which includes harmonised BHPS waves, are discussed in order to assist researchers in identifying what datasets are available and how they can be used.

To conduct research at the geographical level using UKHLS datasets, there are three options:

Firstly, all three of the main Understanding Society datasets (SN 6614, SN 6931 and SN 6676) contain the Government Office Region (GOR) variable **w_gor_dv** which allow the country (England, Scotland, Northern Ireland and Wales) of each household to be identified and in the case of England, the region too. The two-category urban-rural indicator, **w_urban_dv** is also available in all three allowing researchers to identify the household's general urban or rural environment.

Secondly, researchers can apply for datasets containing geographical identifiers which as well as identifying the appropriate geographical area also enables onward linkage of the UKHLS data to external data sources. Each of these datasets contain at least two variables, the household identifier (**w_hidp**) for linkage to other Understanding Society data files and the geographical identifier code itself (see Section 1.1.1 for a full list).

Thirdly, researchers can apply for Understanding Society geographical classification type datasets which although they are not actual geographical identifiers they detail the characteristics of geographical areas. Examples are detailed urban-rural indicators, Output Area Classifications (OAC's) and Acorn. Each of these datasets contain at least two variables, the household identifier (**w_hidp**) for linkage to other Understanding Society data files and the geographical classification itself. (See Section 1.1.2 for a full list).

Note these datasets are available under different access conditions and restrictions; these are outlined in Section 5, Data Access.

1.1 List of Understanding Society geographical datasets

1.1.1 Geographical identifier datasets

All of the following datasets include the wave-specific household identifier (**w_hidp**) for each wave and one or more geographical locator variables as identified in the filename:

- Special Licence Access, Local Authority District (SN6666)
- Special Licence Access, Westminster Parliamentary Constituencies (SN 6668)
- Special Licence Access, Local Education Authorities (SN 6671)
- Special Licence Access, Travel to Work Areas (SN 6675)
- Special Licence Access, Strategic Health Authorities (SN 6672)
- Special Licence Access, Primary Care Organisations (SN 6673)
- Special Licence Access, Census Area Statistics Wards (SN 6669)

- Special Licence Access, Census 2001 Middle Layer Super Output Areas (SN 7245)
- Special Licence Access, Census 2011 Middle Layer Super Output Areas (SN 7249)
- Special Licence Access, Census 2021 Middle Layer Super Output Areas (SN 9170)
- Special Licence Access, Census 2001 Lower Layer Super Output Areas (SN 6670)
- Special Licence Access, Census 2011 Lower Layer Super Output Areas (SN 7248)
- Special Licence Access, Census 2021 Lower Layer Super Output Areas (SN 9169)
- Secure Access version of Understanding Society (which includes Grid References) (SN 6676)

As well as identifying specific geographical locations these geographical identifiers can be used to enrich research by linking the study with external data (either the researchers own or those provided by a third-party organisation). A list of open-access data sources that can be linked is provided In Section 5.1.

1.1.2 Geographical classification datasets

All of the following datasets include the wave-specific household identifier (**w_hidp**) for each wave and one or more geographical classification type variables as identified in the filename:

- Special Licence Access, Census 2001 Rural-Urban Indicators (SN 7454)
- Special Licence Access, Census 2011 Rural-Urban Indicators (SN 7630)
- Special Licence Access, Acorn Type 2015 and 2023 (SN 7453)
- Special Licence Access, Wellbeing Acorn (SN 9385)
- Special Licence Access, Census 2001 Output Area Classification (SN 6674)
- Special Licence Access, Census 2011 Output Area Classification (SN 7629)
- Special Licence Access, Geographical Accessibility (SN 7533)

2. Geographical identifiers in Understanding Society

Geographical identifiers identify geographical areas commonly used in statistical reporting or research. In the UK there are seven main groups of geographical identifiers. Six of these groups are available in UKHLS datasets:

1. Administrative geographies
2. Statistical geographies
3. Census geographies
4. Electoral geographies
5. Health geographies
6. Postal geographies

Figure 1 below illustrates the complexity and multi-layered nature of UK geographies. It presents a hierarchical and diverse system that varies depending on the country (England, Wales, Scotland, Northern Ireland) or the specific type of geography, such as administrative geographies. For a comprehensive overview of UK geographies, see 'A Beginner's Guide to UK Geography' (ONS, 2023).

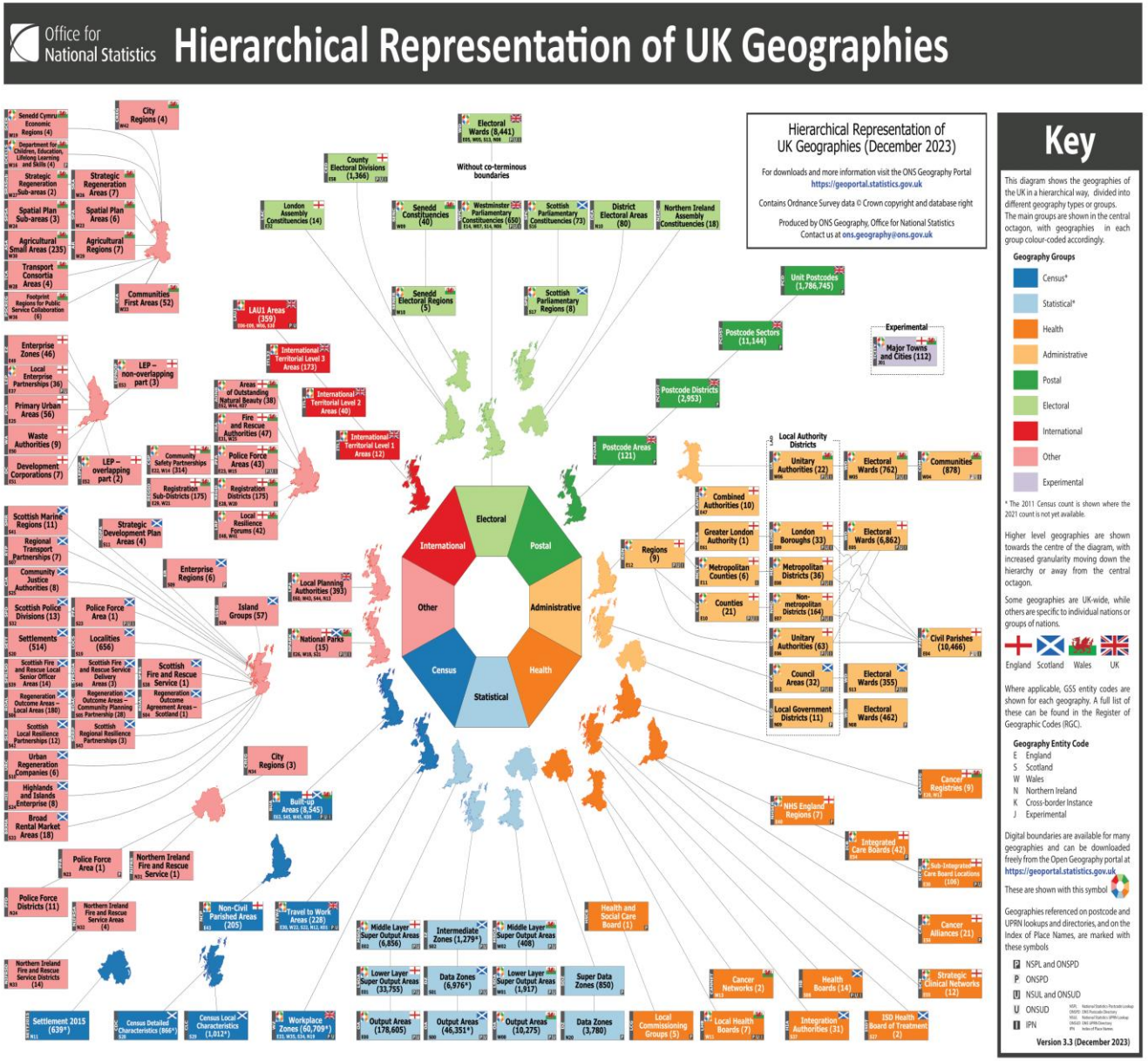


Figure 1: Hierarchical representation of UK Geographies. Source: The Open Geography portal (ONS, 2023).

2.1 Administrative geographies

Like many countries, the UK has its own hierarchy of geographical areas, shaped by the structure of national and local government (ONS, 2023). This hierarchy defines the rights and responsibilities of various administrative units. For example, units can apply to education, transport or policing (e.g., counties in England), while others manage local planning, housing, and refuse collection (e.g. local authority districts/unitary authorities in England) (ONS, 2023).

The number and classification of these units vary between different UK countries, as shown in **Figure 2** below.

Note: The letter in brackets denotes one of the four UK countries (England, Wales, Northern Ireland, Scotland) followed by two digits indicating the level of geography.

Figure 2: Subregional Division of Administrative Units in the UK. Source: A Beginner's Guide to UK Geography (ONS, 2023).

Whilst the administrative division of local units differ considerably across the four UK countries, all four share a common unit: electoral wards (discussed in section 2.b). The primary purpose of these wards is to ensure voting equality throughout the UK, but they also act as foundational spatial units for administrative divisions across countries. The example below shows the spatial division of administrative units. It illustrates the administrative hierarchy in one of the nine English regions – East of England (often referred

to as the East). In **Figure 3**, the region is divided into multiple counties (e.g., Essex). In **Figure 4**, these counties are further divided into local authority districts and unitary authorities, including non-metropolitan districts.

Figure 3

Figure 4

Figures 3 and 4: Subregional Division of Administrative Units in the UK. Source: A Beginner's Guide to UK Geography (ONS, 2023).

For more information about different administrative geographies see ONS Guide to ONS Geography Postcode Products and the already mentioned Beginner's Guide to UK Geography. The maps used in this user guide and other interactive maps are available on the Open Geography portal.

Note: While Wales, Scotland, and Northern Ireland don't possess equivalents to the English

regions and aren't classified as regions, statistical data concerning these countries is frequently included in regional datasets.

Understanding Society provides information on the Regions (formerly Government Office Regions and Local Authority Districts that the households are located in. These Regions are available in the main survey datasets, e.g. SN 6614 and the Local Authority information is available in Special Licence Access, Local Authority District (SN 6666).

The dataset, Special Licence Access, Local Authority District (SN 6666), includes two folders, one for BHPS and one for UKHLS. Within each folder there are data files, one per wave, which are named **w_oslaua_protect** where **w** is the wave prefix. Each datafile includes three variables – the household identifier **w_hidp,** the Local Authority district, **w_oslaua** and the county, **w_oscty**.

a. Statistical and Census geographies

Since the Census occurs once a decade, population fluctuations must be accounted for. To ensure stability in unit sizes and to accommodate population changes between Censuses, statistical geographies were introduced which are small area divisions designed to ensure that units have similar population and household sizes within countries. Unlike administrative geographies, it is intended that the number of occurrences within these units remain consistent over time.

The smallest unit of statistical geographies are Output Areas (OAs) which were established as the core unit for disseminating Census data and was first introduced in England and Wales in 2001, in Scotland in 1984 and in Northern Ireland in 2001. They represent the smallest geographical areas that census data report on. Each country sets its own threshold for how many households make up one OA (ONS, 2023). The number of households in OAs ranges from X to Y in England, A to B in Scotland, C to D in Wales, and E to F in Northern Ireland. It should be noted that Understanding Society does not release geographical information at the Output Area (OA) level.

The next two levels of statistical geographies are Lower Layer Super Output Areas (LSOAs) and Middle Layer Super Output Areas (MSOAs), created by aggregating Output Areas into LSOAs or MSOAs, to ensure that populations within these areas remain within established thresholds (see Table 2 below) as they change over time (ONS, 2023).

For example, LSOAs might be divided into smaller units or multiple LSOAs might be merged into one. The precise thresholds defining LSOAs and MSOAs, how they have changed over time, and the number created in each Census can be found in the ONS beginner's guide.

While they follow a similar rationale, there are subtle differences in the labelling of LSOAs and MSOAs in Scotland and Northern Ireland. In Scotland, the equivalents are Data Zones and Intermediate Data Zones, whereas Northern Ireland uses Super Output Areas, with no direct counterpart for MSOAs. **Note** For the 2021 Census, Northern Ireland introduced new geographical levels: Data Zones and Super Data Zones, replacing the previous Super Output Areas.

Table 2: Population and household minimum and maximum thresholds for OAs, LSOAs and MSOAs.

Geography	Minimum population	Maximum population	Minimum number of households	Maximum number of households
OAs	100	625	40	250
LSOAs	1,000	3,000	400	1,200
MSOAs	5,000	15,000	2,000	6,000

Source: A Beginner's Guide to UK Geography (ONS, 2023).

In the 2001 Census, a special statistical area, the 2001 Census Area Statistic (CAS) ward, was introduced by aggregating Output Areas to produce census statistics, making them comparable to Electoral Wards. These 1,222 aggregations fall within a council area boundary and include approximately 20 households and 50 persons. For more information see https://www.data.gov.uk/dataset/c4afbf66-5d7c-4c41-a52c-9834116841cd/census-2001-cas-wards

In Understanding Society, CAS wards, LSOAs, MSOAs are available as follows but users should note that they cannot request access to geographical datasets from multiple censuses together, for example LSOA2001 and LSOA2011 as the combination can lead to the identification of small areas due to boundary changes between the censuses.

- Special Licence Access, Census Area Statistics Wards (SN 6669)
- Special Licence Access, Census 2001 Middle Layer Super Output Areas (SN 7245)
- Special Licence Access, Census 2011 Middle Layer Super Output Areas (SN 7249)
- Special Licence Access, Census 2021 Middle Layer Super Output Areas (SN 9170)
- Special Licence Access, Census 2001 Lower Layer Super Output Areas (SN 6670)
- Special Licence Access, Census 2011 Lower Layer Super Output Areas (SN 7248)
- Special Licence Access, Census 2021 Lower Layer Super Output Areas (SN 9169)

These datasets include two folders, one for BHPS and one for UKHLS (except for SN 9170 and SN 9169 which only include the UKHLS folder). Within each folder, there are data files, one per wave, which are named **w_geoYR** where **w** is the wave prefix, **geo** is either lsoa or msoa and **YR** is either 01, 11 or 21 depending on the Census it refers to. Each data file includes two variables – the household identifier **w_hidp** and the geographical identifier **w_pcon_protect**

Note In Understanding Society the Output Area Classifications that are provided are not these statistical geographies but geodemographic areas, see Section 3.c for further details. Understanding Society also provides rural-urban indicators for census data, see Section 3.a for further details.

b. Electoral geographies

The UK has a complex and evolving electoral system that includes elections for the UK parliament, devolved governments (Wales, Scotland, and Northern Ireland), county councils and local authorities. There are 650 UK parliamentary constituencies also known as Westminster Parliamentary Constituencies (WPC)s in the UK: 543 in England, 57 in Scotland, 32 in Wales, and 18 in Northern Ireland.

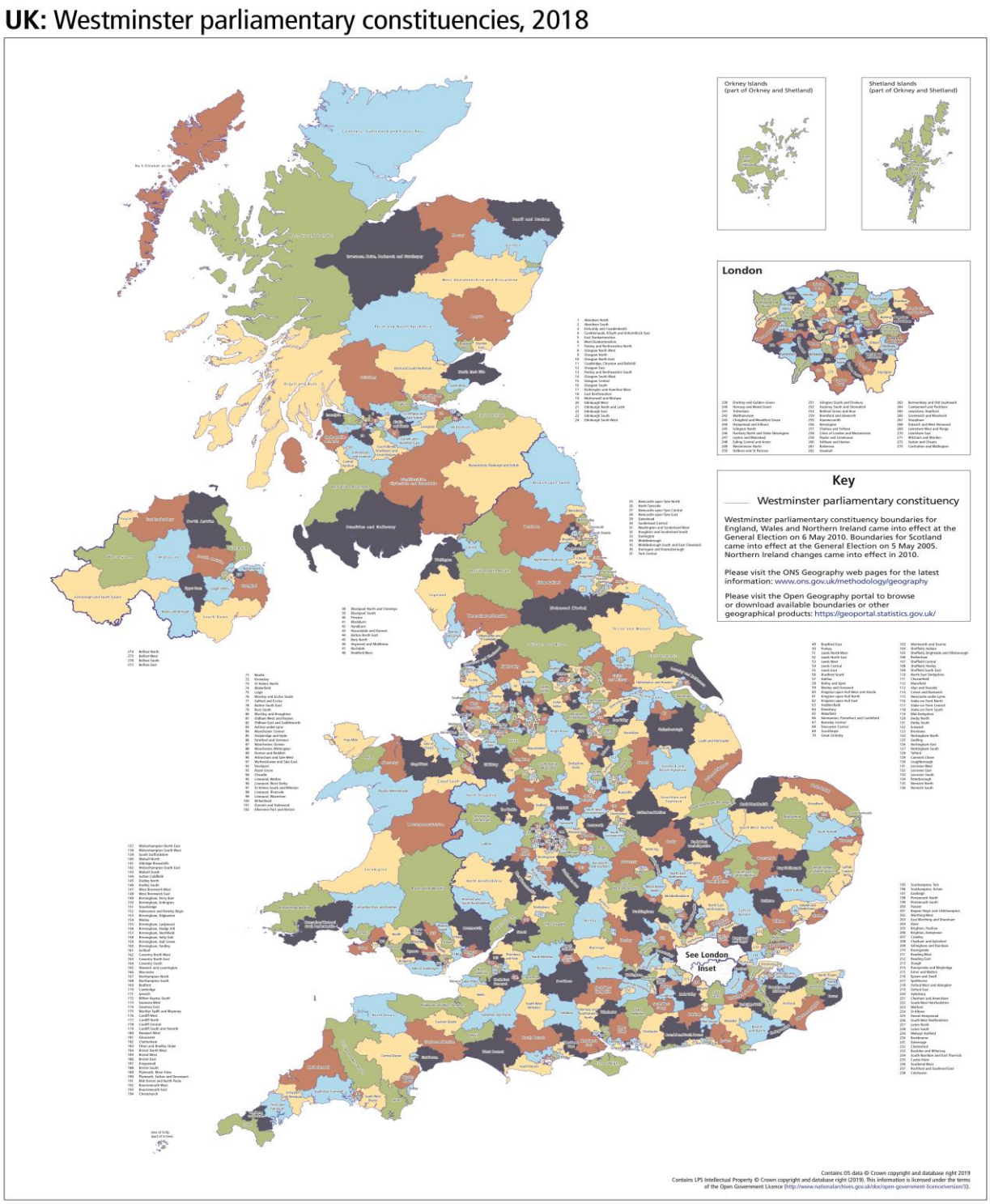


Figure 5: Westminster parliamentary constituencies (2018). Source: The Open Geography portal (ONS, 2023)

Before each election, a review of the Westminster constituency sizes is conducted to ensure equality of votes, which may result in changes to the constituency boundaries or even the removal or introduction of constituencies. It is important to note that constituencies are usually based on Local Authority Districts (LADs) / Unitary Authorities (either fully or partially), unless there is a strong case for a different arrangement (ONS, 2023).

Understanding Society provides geographical identifiers for these UK parliamentary constituencies (WPCs), Special Licence Access, Westminster Parliamentary Constituencies (SN 6668)

The Special Licence Access, Westminster Parliamentary Constituencies (SN 6668) dataset includes two folders, one for BHPS and one for UKHLS. Within each folder there are data files, one per wave, which are named **w_pcon_protect** where **w** is the wave prefix, and each data file includes two variables – the household identifier **w_hidp** and the geographical identifier **w_pcon_protect**.

c. Health geographies

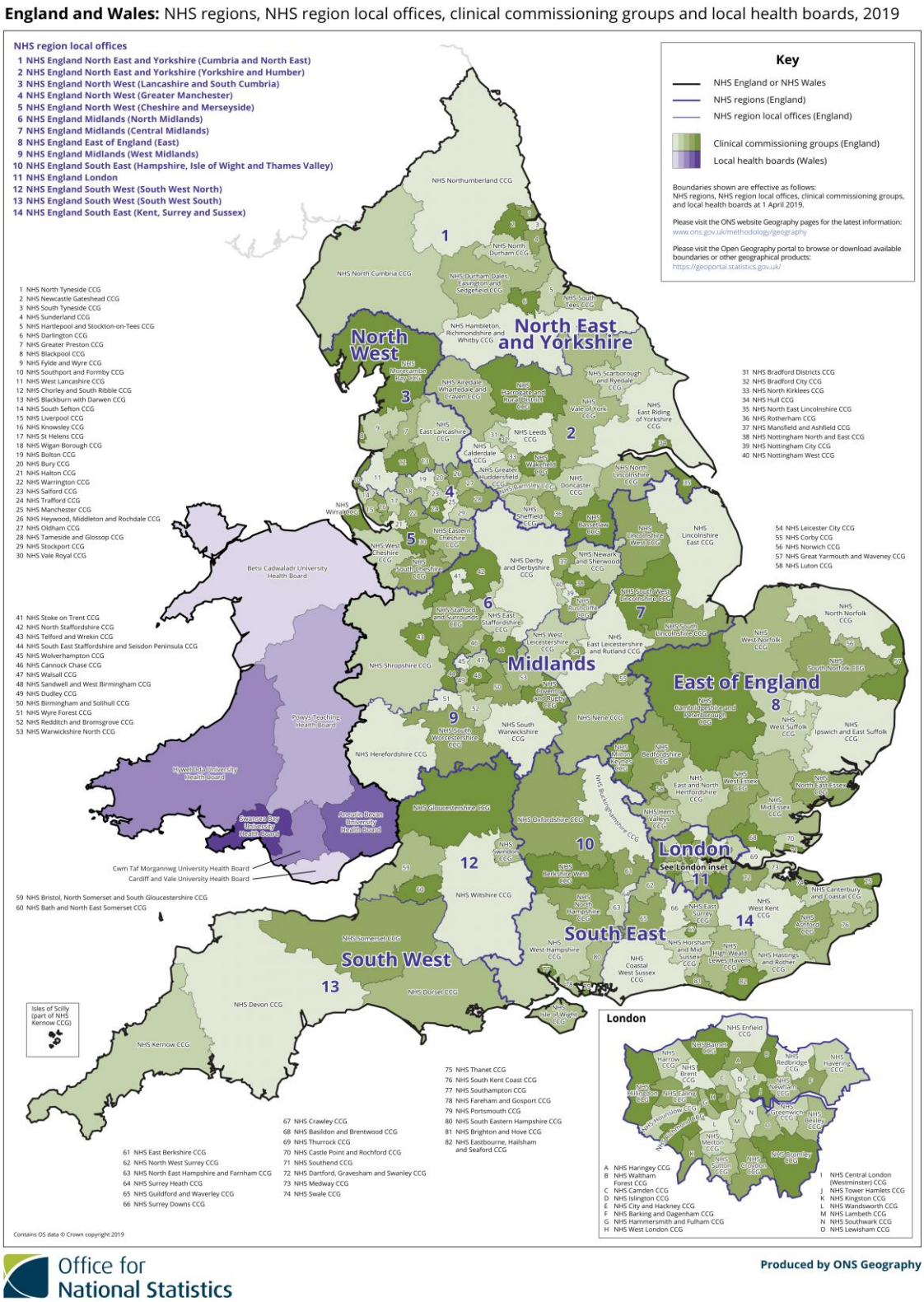


Figure 6: Health area division in England and Wales (2019). Source: The Open Geography portal (ONS, 2023).

Health authorities are structured to align with the organization of the NHS, which is distinct

for each constituent country of the UK, reflecting their unique structures. The Understanding Society data includes two specific health geographical identifiers: Strategic Health Authorities (SHA) and Primary Care Organisations (PCO). These identifiers represent the geographical divisions within the UK's healthcare system and include the equivalent entities for Wales, Scotland, and Northern Ireland. For a more comprehensive understanding of the various health geographies and their evolution, refer to the ONS beginner's guide (ONS, 2023). **Note:** The structure of Strategic Health Authorities was significantly reorganised in April 2013.

Understanding Society provides Strategic Health Authority identifiers, corresponding to the former configuration used between October 1st, 2002 to March 31st, 2013. Primary Care Organisation identifiers are also provided.

- Special Licence Access, Strategic Health Authorities (SN 6672)
- Special Licence Access, Primary Care Organisations (SN 6673)

The dataset Special Licence Access, Strategic Health Authorities (SN 6672), includes two folders one for BHPS and one for UKHLS. Within each folder there are data files, one per wave, which are named **w_oshlthau** where **w** is the wave prefix, and each data file includes two variables – the household identifier **w_hidp** and the geographical identifier **w_oshlthau**.

The dataset Special Licence Access, Strategic Health Authorities (SN 6673), includes two folders, one for BHPS and one for UKHLS. Within each folder, there are data files, one per wave, which are named **w_pct** where **w** is the wave prefix, and each data file includes two variables – the household identifier **w_hidp** and the geographical identifier **w_pct**.

d. Postal geographies

Royal Mail operates a comprehensive network of postcodes to define postal delivery zones across the UK. Since most individuals are familiar with their postcode this information is collected from respondents and used by the Understanding Society team internally as the basis for identifying all types of geographies.

In Understanding Society postcodes are not released, however, grid references of postcode centroids are available in the Secure Access version of the main study dataset.

- Secure Access version of Understanding Society (which includes Grid References) (SN 6676)

e. Other geographies

In the UK, geographical divisions go beyond the usual administrative, statistical (census), and electoral boundaries. There is a diverse array of geographical identifiers, each designed to meet the specific needs of different territorial segmentations in various governmental domains. This system includes local education authorities, travel-to-work areas, and more. The following section elaborates on these geographical identifiers, focusing on those included in Understanding Society data.

2.1.1 Local Education Authorities

Local Education Authorities (LEAs) served as the administrative agencies for educational services in England. Up until 2009 each county, unitary authority, metropolitan district, and London borough had an LEA. Despite the discontinuation of LEAs in 2009, the succeeding educational authority divisions still respect the geographic demarcations set by the former LEAs. The ONS beginner's guide (ONS, 2023) provides detailed insights into the new coding system and corresponding structures in other UK countries.

In Understanding Society, Local Education Authorities / Education and Library Boards geographical identifiers are provided:

- Special Licence Access, Local Education Authorities (SN 6671).

This dataset includes two folders, one for BHPS and one for UKHLS. Within each folder, there are data files, one per wave, which are named **w_ oslaua** where **w** is the wave prefix, and each data file includes two variables – the household identifier **w_hidp** and the geographical identifier **w_ oslaua**.

2.1.2 Travel To Work Areas

Travel To Work Areas (TTWAs) are designed to aid in examining the labour market and associated planning activities. Central to this examination are commuting patterns (ONS, 2023). The ONS defines a TTWA as a region where "at least 75% of the residents' workforce is employed within the area, and at least 75% of the employees in the area live there" (ONS, 2023: 56). For more detailed information on the structure of TTWAs within the various UK countries, refer to The ONS Beginner's Guide (ONS, 2023).

In Understanding Society, TTWA geographical identifiers as defined for both the 2007 and 2015 variants are provided:

- Special Licence Access, Travel to Work Areas (SN 6675).

This dataset includes two folders, one for BHPS and one for UKHLS. Within each folder there are data files, one per wave, which are named **w_ ttwa07 and w_ttwa15** where **w** is the wave prefix, and each data file includes two variables – the household identifier **w_hidp** and the geographical identifier **w_ ttwa07 amd w+ttwa15**.

3. Geographical classification datasets

a. Urban-Rural classifications

The urban-rural classifications are produced by the ONS for England and Wales, the Scottish Government for Scotland, and NISRA for Northern Ireland, for each Census.

The two-category rural / urban categorisation variable is available in the main survey datasets (SN 6614, SN 6931 and SN 6676) whilst the full categories are available as follows:

- Special Licence Access, Census 2001 Rural-Urban Indicators (SN 7454)
- Special Licence Access, Census 2011 Rural-Urban Indicators (SN 7630)

The Special Licence Access, Census 2001 Rural-Urban Indicators (SN 7454), dataset contains two folders, one for BHPS and one for UKHLS. Within each folder there are data files, one per wave. These data files are named **w_ ur01ind_protect** which includes the rural-urban categorisation variable, **w_ur01ind** and the household identifier, **w_hidp**. The value labels for these urban rural classification variables are available from the ONS for England and Wales.

The urban-rural division reflects differences in population density at the census level, classifying any area with more than 10,000 inhabitants as urban. A binary identifier, **urban_dv**, classifying the address as falling into an (1) urban or (2) rural area is derived from the Office for National Statistics Rural and Urban Classification of Output Areas 2001 (UK Data Service Study Number 7454). The identifier assumes a value of (1) if the address falls within urban settlements with a population of 10,000 or more, or (2) otherwise.

The Special Licence Access, Census 2011 Rural-Urban Indicators (SN 7630) dataset contains one folder, UKHLS, which contains data files, one per wave. These data files are named **n_ru11ind_protect** and include the rural-urban categorisation variable, **w_rur01ind** and the household identifier, **w_hidp**. These classifications are not available for Northern Ireland. The value labels for these urban rural classification variables are available from the ONS for England and Wales.

w_ru11ind	
England and Wales	
A1	Urban major conurbation (only for England)
B1	Urban minor conurbation (only for England)
C1	Urban city and town
C2	Urban city and town in a sparse setting
D1	Rural town and fringe
D2	Rural town and fringe in a sparse setting
E1	Rural village
E2	Rural village in a sparse setting
F1	Rural hamlet and isolated dwellings
F2	Rural hamlet and isolated dwellings in a sparse setting

-9	*Address postcode missing*
Scotland	
1	Large Urban Area
2	Other Urban Area
3	Accessible Small Town
4	Remote Small Town
5	Very Remote Small Town
6	Accessible Rural
7	Remote Rural
-9	*Address postcode missing*
Northern Ireland	
-9	Missing

b. Acorn geographical identifiers

Understanding Society provides Acorn geographical identifiers which contains geodemographic segmentation of residential neighbourhoods in the UK. It classifies each postcode into one of 64 types, offering a distinctive picture of the people who typically live in each area, their attitudes, and their behaviours. Acorn types are developed by CACI using data from various sources (including Understanding Society), incorporating new data as it becomes available. In 2023, the Acorn classification was completely rebuilt to create new segmentation, with updated categories, groups, and types that reflect recent social, demographic, and economic changes. For more information on this classification please visit the CACI website and documentation section on the UK Data Service website.

In Understanding Society, Acorn geographical categories are provided in:

- Special Licence Access, Acorn Type 2015 and 2023 (SN 7453).

This dataset contains two sub-folders one for BHPS and another for UKHLS. Each of these contain data files, one per wave, which are named **w_acorn_protect** where w is the wave prefix. Each data file includes Acorn geodemographic segmentation codes **(w_acorncategory2015, w_acorngroup2015, w_acorntype2015, w_acorncategory2023 w_acorngroup2023, and w_acorntype2023**) for each household in the study, along with a household identifier, **w_hidp**, allowing it to be linked to the main Understanding Society data files.

In addition to the standard Acorn geodemographic segmentation, Understanding Society also supplies Special Licence Access, Wellbeing Acorn (SN 9385) a health- and lifestyle-focused segmentation created by CACI. Wellbeing Acorn assigns UK postcodes to ranked types and groups that summarise health behaviours and circumstances. It follows the same per-postcode principle as Acorn and is designed to be used alongside main survey data.

Note that this dataset includes only UKHLS waves. For details of the classification, see the CACI website and the dataset documentation.

c. Output Area Classifications

The Output Area Classification (2021 OAC) is a Census hierarchical geodemographic system used across the UK to identify areas with similar characteristics. Developed collaboratively by the Office for National Statistics (ONS) and University College London (UCL), the classification organises Output Areas (OAs) into 8 supergroups, which are further divided into 21 groups and 52 subgroups (CDRC, 2025).

As an example the Retired Professionals supergroup is broken down into Spacious Rural, Small Town Suburbia, or Established Mature Families group. These three groups are further divided into Pre-Retirement Spacious Living and retirement Spacious Living (Spacious Rural Living group); Younger Established Suburban Communities and Older Established Suburban Communities (Small Town Suburbia group); Affluent Mature Families and Burgeoning Mature Families (Established Mature Families group).

The OAC classifications are provided as follows:

- Special Licence Access, Census 2001 Output Area Classification (SN 6674)
- Special Licence Access, Census 2011 Output Area Classification (SN 7629)

These datasets contain two sub-folders one for BHPS and another for UKHLS. Each of these contain data files, one per wave, which are named **w_oacYR_protect** where w is the wave prefix and YR is 01 or 11 depending on the Census year it refers to. Each data file includes the OAC variables, w_oac01 **w**_oac01spgrp **w**_oac01grp, w_oac01sbgrp, and the household identifier, w_hidp, which allows linkage to the main data files. Value labels are available in the documentation.

d. Accessibility dataset

The Understanding Society Geographical Accessibility dataset (SN 7533) contains information from the Department for Transport (DfT)'s Accessibility Statistics linked with information from the first three waves of the UKHLS. It provides information about the areas in which study members live, taken from more than 20 published tables describing small areas in England (N=32,484). They contain more than 600 unique data items relating to access to eight domains of public service (i.e., Employment Centres, Primary Schools, Secondary Schools, Further Education, General Practitioners, Hospitals, Food Stores, and Town Centres) (University of Essex et al., 2024: 3).

Information about the dataset, including the comprehensive User Guide explaining how the linkage between the UKHLS data and DfT data was made can be found in the dataset's documentation.

4. Derivation Methodology

Although not released, the postcode for every household in Understanding Society is held by ISER. Those postcodes are used as the basis for producing the various geographical data files that are detailed in this document.

With the exception of the Acorn datasets all other geographical identifier and classification variables are extracted from ONS Postcode Directories (ONSPD) by matching the postcode of the household with the postcode in the relevant ONSPD.

ONSPD's are publicly available from ONS and are released four times per year. The default ONSPD used for each Understanding Society wave is the one that is closest to the mid-point of the fieldwork phase for that wave. The May 2023 version of the ONSPD is used for Wave 14 of Understanding Society for example. For some geographical variables, however, other ONSPDs have to be used, typically because the ONSPD for the mid-point of a specific wave doesn't contain the information required. An example is the Census 2011 LSOA variable which wasn't available until the May 2015 version of the ONSPD. Consequently, for Waves 1 to 5 they were extracted from the May 2015 version instead of the default ONSPD for those waves. Similarly, some geographical identifiers ceased to be released in ONSPD's so earlier versions of them have to be used.

It should be noted that due to the lack of ONSPDs for many BHPS waves the November 2013 version has been used to derive all geographies for the harmonised BHPS waves.

As noted above the Acorn datasets are not derived from ONSPDs but from lookup files supplied directly to the Understanding Society team, again at the postcode level.

A separate document is released with each wave of Understanding Society that contains additional information on the makeup of the geographical variables. The information includes:

- A table identifying the ONSPD used per variable per Wave
- Details on the methodology for deriving the Acorn datasets
- A table detailing the numbers of matched and missing / unmatched households to postcodes
- A table containing a full list of geographical variable names together with their Stata data type and variable label

The document can be found in the Documentation section of every Understanding Society geographical Study Number (SN) on the UK Data Service data catalogue, the links for which are provided in the relevant sections of this document.

5. Data Access

The geographical identifier and classification datasets are rigorously anonymised to comply with exacting security measures, yet they encompass a multitude of variables that could potentially risk disclosure when combined. For instance, while the identification of a location at the local authority level in isolation may be low risk, combining this with additional data points might elevate the risk of personal data exposure. To safeguard participant privacy while enabling as much research as possible, our data repository provider, the UK Data Service (UKDS), applies three progressive levels of access as agreed with the Understanding Society team:

1. Safeguarded - End User Licence (EUL)
2. Safeguarded - Special Licence (SL)
3. Controlled - Secure Access (SA)

The levels reflect the varying degrees of sensitivity within the geographical identifier and classification datasets—the more sensitive the information, the more restricted the access. Data under EUL have the lowest perceived risk of disclosure, whereas SA governs the most sensitive datasets. **Note** that none of the Understanding Society geographical identifier and classification datasets are available under Open Access, thus a licence agreement is mandatory for access in every case.

To assist users in navigating the access requirements for our geographical identifier and classification datasets and to facilitate a seamless application process, the Understanding Society team has worked with the UK Data Service to develop a variety of supportive resources. These resources include detailed guides, webinars, and other materials aimed at guiding users through the application process required to access Understanding Society geographical data. Further details about these resources, along with broader information concerning our data, are available on the Understanding Society website.

5.1 Linking Understanding Society to open data sources

A list of open data sources that can be linked to Understanding Society is available on the Understanding Society website. This list is not exhaustive and will be updated periodically.

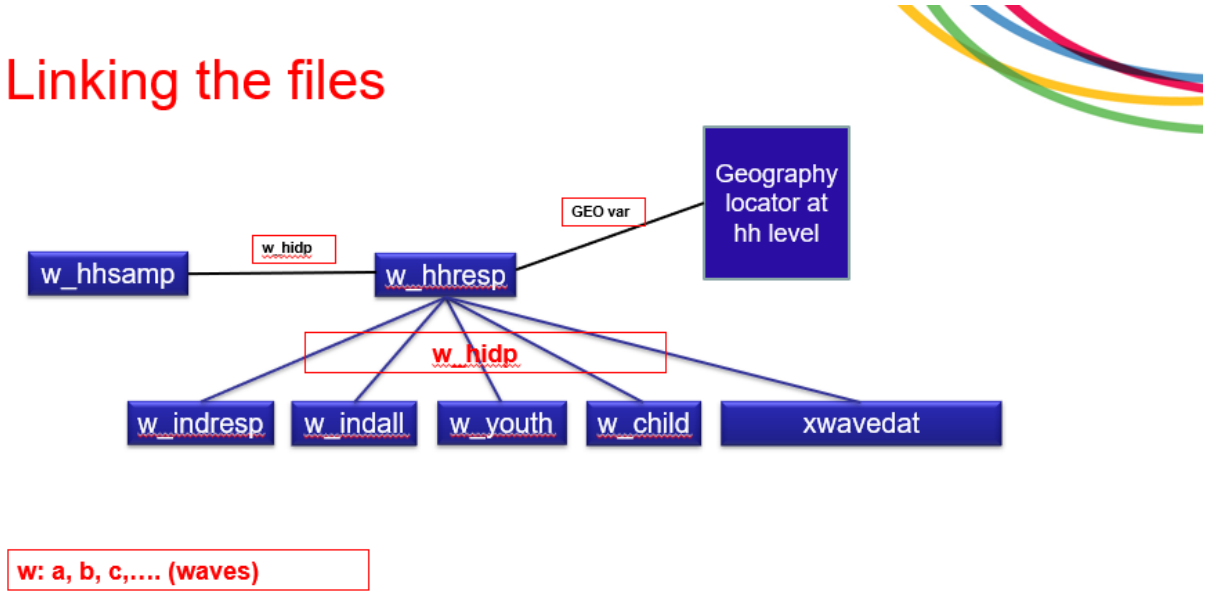
5.2 How to link data

Once access to the desired geographical dataset(s) has been granted, researchers will receive a dataset containing a file for each wave. These files will contain at least two variables: the household identifier (**hidp**) and the geographical identifier(s), formatted as a 9-character string (with a letter denoting the country), e.g.:

w_hidp	LAD23CD
1	E06000020
2	N09000003
3	W06000013
4	S12000041

In this case, the linking variable is **LAD23CD**, a geographical identifier found in both the Understanding Society datasets and the external dataset being linked to.

To link the Understanding Society geographical dataset to other Understanding Society datasets, follow these steps. The household identifier, **w_hidp**, should first be linked to the household response file (**hhresp**). This can then be connected to the rest of the UKHLS datasets as follows:



5.3 Potential challenges when linking the data to external data sources

As society changes over time the geographies describing it change too. Therefore, it is important to be aware of the following when linking external data to UKHLS data, using geographical identifiers:

- Matching timelines – data collection (fieldwork schedule) and the seasonality can vary between datasets, e.g. differences between calendar and financial year, data being collected monthly, quarterly, yearly.

- Terminology – when exploring open sources pay attention to what information is provided about available geographies, e.g. 'local authorities'. Sometimes different data sources will refer to the same unit in different ways – local authorities, local authority districts, non-metropolitan districts, etc.

- Changes over time – some administrative units change their size over time, e.g. change in the number of districts in the county of Somerset.

- While the Understanding Society sample is representative at each geographical level, small sample sizes in some areas can make analysis more restrictive and challenging. This limitation may lead to decreased generalizability of the results, increased margins of error, and other caveats that affect the reliability of the findings.

- While Understanding Society provides Lookup tables to show which year of the ONS postcode directory our geographical identifiers were extracted from the, the same information is not often available in external data sources. Additionally, not all geographical identifiers follow the same chronological pattern (some are irregular), so be cautious when conducting longitudinal analysis with various geographical levels.

6. Should we include other geographical identifiers?

Since UK geographies offer a rich array of identifiers not fully covered in this user guide, we invite you to help enhance the list of available geographical identifiers in Understanding Society. If you have suggestions for geographies we should include in future waves of our study, please email us at User Support.

References

Consumer Data Research Centre. (2023). Output Area Classification (2021). Retrieved from https://data.cdrc.ac.uk/dataset/output-area-classification-2021

Office for National Statistics. (2023). A beginner's guide to UK geography: Geographical areas frequently used in the publication of official statistics (August 2023 v2.0). Retrieved from https://geoportal.statistics.gov.uk/

Office for National Statistics. (2023). Geoportal. Retrieved from https://geoportal.statistics.gov.uk/

UK Data Service. (2025). Retrieved September 22, 2025, from https://ukdataservice.ac.uk/

University of Essex, Institute for Social and Economic Research, NatCen Social Research, Kantar Public. (2025). *Understanding Society: Waves 1-2, 2009-2011: Special Licence Access, Geographical Accessibility*. [data collection]. *2nd Edition.* UK Data Service. SN: 7533, DOI: http://doi.org/10.5255/UKDA-SN-7533-2

University of Essex, Institute for Social and Economic Research. (2025). Understanding Society - Linked Data. Retrieved from https://www.understandingsociety.ac.uk/documentation/linked-data