# Admin-based statistics for property type, feasibility research: England and Wales

Early research demonstrating the potential of Valuation Office Agency data to provide detailed information on property type, how it compares to 2011 Census accommodation type, and a request for feedback on the usefulness of these statistics.

Contact:
Lizzie Gunner, Alistair Dent, Stephan Tietz
admin.based.
characteristics@ons.gov.uk
+44 (0) 1329 444528

Release date:
10 March 2022

Next release:
To be announced

## Notice

### 10 March 2022

The latest article in this series is located in a different area of the website.

# Table of contents

# 1 . Disclaimer

These Research Outputs and the linked Valuation Office Agency (VOA) property details data used for the analyses are not official statistics on accommodation types[1] in England and Wales. Rather they are published to allow users to compare the VOA data with the 2011 Census, as a potential alternative approach to collecting information on the accommodation type. These outputs should not be used for policy- or decision-making. Official statistics about property type include Council Tax statistics and 2011 Census accommodation type.

It is important the information and research presented are read alongside the analysis to aid interpretation and avoid misunderstanding. These analyses must not be reproduced without this disclaimer and warning note.

**Notes for: Disclaimer**

1. Accommodation type is a measure of the property. For example, flat, terraced house and so on.

# 2 . Summary

We are exploring the use of administrative data on housing as a replacement for collecting this information in censuses and surveys. This research compares property type data recorded by the Valuation Office Agency (VOA) with data collected by the 2011 Census in England and Wales.

This is part of our ongoing research exploring how VOA data could be used to produce accommodation statistics. This follows up on our commitment to explore the feasibility of supplementing census with VOA data on property type, size and rooms as stated in the White Paper, Help Shape Our Future: The 2021 Census of Population and Housing in England and Wales (PDF, 967KB). We previously published a feasibility report estimating the number of rooms using data recorded by the VOA.

Administrative data have the potential to remove the need for property type data to be collected via a census or surveys and provide new breakdowns of property types not available from a census or surveys. This may help meet user needs for more detailed categories of property type. For example, new analysis could include being able to distinguish between bungalows, houses and different types of terraced houses.

For this work we successfully linked 96% of 2011 Census households[1] to VOA property data. We found exact agreement between the VOA property type variable and the 2011 Census accommodation type variable for 86% of linked addresses. This rose to 93% when more ambiguous matches (which we will refer to as "explainable agreement") between census and VOA property type were included. This suggests there is potential to produce new property type statistics from such data, and our course of action will depend on feedback on this from users. More information on agreement categories can be found in Section 6.

Further research is required to establish if administrative data can provide comparable information with the 2011 Census category for dwellings "above or within commercial building" and to understand how to address the small proportion of records where VOA property type and census accommodation type appear contradictory.

**Notes for: Introduction**

1. Census households excluding non-responding households and those with a non-unique (duplicate) unique property reference number.

# 3 . Things you need to know about this release

We are transforming the way we produce population, migration and social statistics to better meet the needs of our users and to produce the best statistics from all the available data. More information about our plans to do this and how we are progressing a programme of work to put administrative data at the core of population, migration and social statistics is available.

We welcome feedback on these Research Outputs on the findings, methodology and whether these alternative outputs would meet your information needs. Please email your feedback to admin.based.characteristics@ons.gov.uk and include "Housing" in the subject line of your response.

This is early research to demonstrate the potential of administrative data to provide information on accommodation type, which is widely collected by surveys and census. In order to compare administrative data with self-reported data, we linked the 2011 Census and data from the Valuation Office Agency (VOA). More information about the VOA data can be found in the source overview.

# 4 . Background

In the 2018 White Paper, Help Shape Our Future: The 2021 Census of Population and Housing in England and Wales (PDF, 967KB), we committed to exploring the feasibility of supplementing 2021 Census questions with administrative data on total number of rooms, living space and property type.

Housing statistics are widely used in local and central government for planning and policy development. The main user requirement is to understand the characteristics of the existing housing stock. This enables local housing policymakers to better meet the future housing needs of local residents.

In the latest census consultation (PDF, 83KB), users indicated that they wanted more detail on property type. More data are available from the Valuation Office Agency (VOA) such as additional breakdowns of flats and separation of bungalows from houses. Note that VOA property type categories do not provide information on dwellings above or within commercial buildings.

# 5 . Statistical quality and coverage of VOA data

The 2011 Census records the accommodation type for each household. The 2011 Census defines a household as "one person living alone, or a group of people (not necessarily related) living at the same address who share cooking facilities and share a living room or sitting room or dining area".

Most residential addresses in England and Wales are used by a single household but we identified that 1% of addresses were duplicated, which may be because there was more than one household at an address. In contrast, the Valuation Office Agency (VOA) data hold information on addresses and it is not currently possible to clearly identify multiple households at an address from administrative data alone. If a property is empty, information would still be captured by the VOA but would not be recorded by the 2011 Census as it would be classed as a non-response.
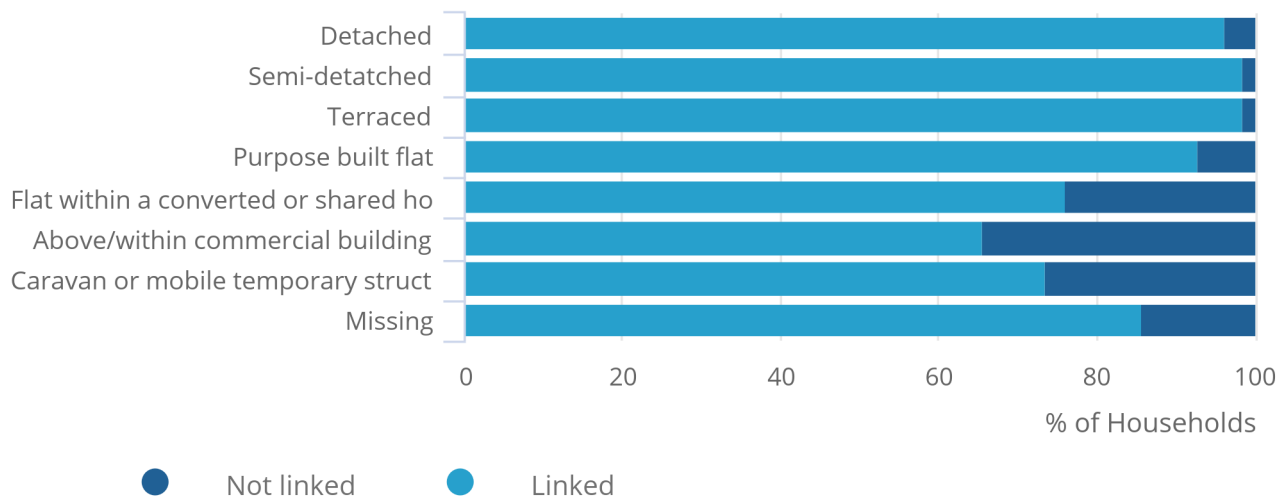
For this analysis a unique property reference number (UPRN) was assigned to records in both datasets, and cases were linked on this identifier. A UPRN is a unique identifier for every address in Great Britain and is allocated by local government and Ordnance Survey. Where we could not assign a UPRN or where there were cases of two or more records with the same UPRN, these cases were removed. A more detailed description of this methodology can be found in Section 12.

The proportion of census households that linked to VOA records varied by census accommodation type. High linkage rates (between 92% and 99%) were achieved for detached, semi-detached and terraced properties as well as purpose-built flats. Conversely, 66% of Census households in the category "Above/within commercial building" linked to VOA records.

The distribution of records that linked for each census accommodation type is shown in Figure 1.

## Figure 1: Proportion of census households that linked to VOA property data



Detached
Semi-detatched
Terraced
Purpose built flat
Flat within a converted or shared ho
Above/within commercial building
Caravan or mobile temporary struct
Missing

0    20    40    60    80    100

% of Households

● Not linked    ● Linked

**Source: Valuation Office Agency, Office for National Statistics © Crown Copyright 2015. This work uses research datasets which may not exactly reproduce National Statistics aggregates**

**Notes:**

1. These are shortened wordings. See Annex 3 for complete wording of the census question categories.

Overall, all census accommodation types had a linkage rate of over 66%. Fewer flats were linked than other accommodation types, which could suggest there might be a small issue with the linkage of flats. If areas have a larger proportion of flats, they may see lower linkage rates. On Figure 1, the missing category is where the householder has not responded to the census accommodation type question or gave an invalid response[1].

The linkage rate overall for households in England and Wales was 96%. The linkage rate for England was 96% and Wales was 94%. Linkage varies across different local authorities ranging from 60% (Isles of Scilly) to 99% (Wyre Forest). Isles of Scilly is an outlier, and the local authority with the second-lowest linkage rate across England and Wales is Kensington and Chelsea (74%).

Although the local authorities at the extremes of the distribution are both in England, note that local authorities in Wales have a similar distribution of linkage rates. In Wales, Ceredigion has the lowest linkage rate (78%) and Merthyr Tydfil has the highest (97%). For all local authorities in England and Wales, 94% have a linkage rate of over 90%. High linkage rates mean that most records can be checked for agreement. All local authorities had less than 5% of missing VOA property type values.

## Notes for: Statistical quality and coverage of VOA data

1. For this research we used 2011 Census data without imputed values.

# 6 . Comparing census accommodation type and VOA property type

The 2011 Census data refer to the types of property (for example, detached house, flat and so on) as "accommodation type" whereas the Valuation Office Agency (VOA) data refer to it as "property type". The categories for each are in Section 14. They are similar concepts with different subcategories, which can be mapped onto each other as shown in Section 13. As these categories differ, we have split agreement level into four categories.

| Agreement category | Meaning |
|---|---|
| Exact agreement | Both variables had the same meaning. For example, "detached" on the census and "detached house" on VOA data. |
| Explainable agreement | Variables did not agree completely but we can explain why this combination did not match but possibly could. For example, "semi-detached house" on the 2011 Census and "end-terraced house" on VOA. |
| More information needed | We do not have enough information to determine if they agree or not. For example, "non- purpose-built flat self-contained" on VOA and "flat, maisonette or apartment that is in or above a commercial building" may be an accurate description of the same property but may not. |
| Contradiction | Variables contradict each other. For example, "purpose-built flat with lift" and "end-terraced house". |

# 7 . Agreement between VOA and 2011 Census

We assigned agreement categories to households for each combination of census accommodation and Valuation Office Agency (VOA) property type (see Section 6). The mapping in Section 13 demonstrates how we might expect VOA property type to match up to 2011 Census accommodation type.

The "exact agreement" rate between the property type on the 2011 Census and the VOA data was 86%. This rises to 93% if "explainable agreements" are included (Figure 2). Exact agreement for England was 86% and for Wales it was 88%.

**Figure 2: Agreement between census accommodation type and VOA property type**

Figure 2: Agreement between census accommodation type and VOA property type
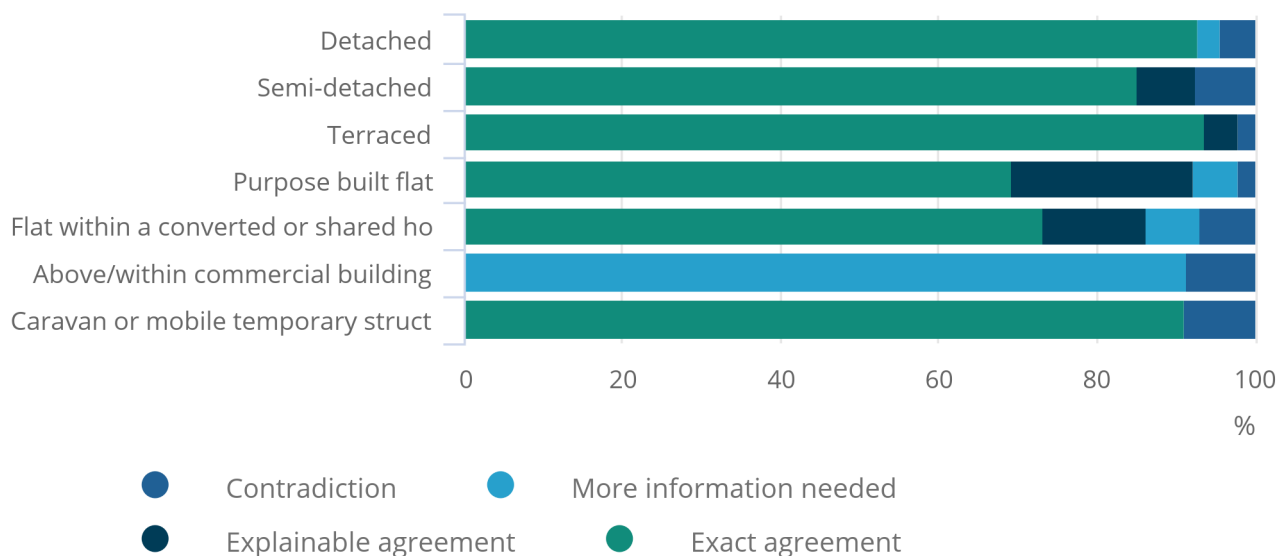


Legend:
- Contradiction
- More information needed
- Explainable agreement
- Exact agreement

**Source: Valuation Office Agency, Office for National Statistics © Crown Copyright 2015. This work uses research datasets which may not exactly reproduce National Statistics aggregates**

The census accommodation types "purpose-built flat" and "flat within a converted or shared house" have a relatively higher proportion of "explainable agreement". The VOA has a category called "flat in two-storey mid-terraced" but the census flat categories are "flat, maisonette or apartment that is in a purpose-built block of flats or tenement" or "flat, maisonette or apartment that is part of a converted or shared house (including bedsits)". The VOA flat types do not all clearly fit into either of these categories and therefore are not an exact agreement.

If a record with the previous VOA property type matched with either of the 2011 Census flat categories, it would fit into the "explainable agreement" agreement category because the VOA category is clearly a flat but whether the VOA category is a "purpose-built flat" or a "flat within a converted or shared house" is unclear." One limitation of VOA for property type is that we cannot accurately know if all flats are purpose-built or converted as only some flats are defined. See Section 13 for more detail on how the categories on the 2011 Census and VOA map to each other.

There are a significant number of records where the property type was "end-terraced house" or "end-terraced bungalow", but the 2011 Census accommodation type was "semi-detached". This can be seen by the proportion of "explainable agreement" for the "semi-detached" 2011 Census accommodation type in Figure 2. This may be because some census respondents have mis-categorised their property as this difference may not be obvious to them, but a trained surveyor may understand this.

The 2011 Census category for "properties within/above a commercial building" has no equivalent VOA property type. This explains why we have no "exact" or "explainable agreement" for this census category. Many of the property types that matched with records with an accommodation type of "A flat, maisonette or apartment that is in a commercial building (for example, in an office building, hotel, or over a shop)" were flats but without a specific variable on the VOA data; we cannot currently identify these properties without a census. VOA collects a variable called "dwelling group code", which might help identify these properties in future as it has more information on properties in commercial buildings. We would like feedback on whether knowing if a property is above or within a commercial building is essential for users.

# 8 . New insights from using VOA data

A potential benefit of using Valuation Office Agency (VOA) property type is that many more categories and breakdowns are available in contrast to data currently collected on most censuses and surveys. The 2011 Census groups flats, maisonettes and apartments together and has three categories for these: "purpose-built", "flat within a converted or shared house" and "above or in a commercial building".

VOA data have 16 categories of flats, which include maisonettes, purpose-built or "non-purpose built", flats with lifts and without lifts and whether the block of flats is terraced, mid-terraced, semi-detached or detached. More detail can be found in Section 13.

Another difference is the separation of bungalows from houses. In the 2011 Census, bungalows are in the same category as houses so it is possible to identify whether a bungalow or house is terraced, detached or semi-detached but not whether it is a bungalow or house. VOA data show houses and bungalows in their own categories. Additionally, houses, bungalows and flats can be broken down by "end of terrace" and "mid-terrace" using VOA data, compared with just "terraced" on the 2011 Census.

VOA property type variable also has separate categories for mobile homes, houseboats and caravans compared with a single category on the 2011 Census.

These extra breakdowns provide the opportunity for more detailed analyses for property type when combined with censuses or other administrative data. Analyses using different breakdowns of property type can be used to get insights on floor space from VOA data.

# 9 . Case studies: local authorities with the highest and lowest direct agreement

We have selected two local authorities: one that had a high "exact agreement" rate and one that had both a low "exact agreement" rate and a high percentage of contradictions.

## Case study for Oadby and Wigston

Oadby and Wigston is a predominantly urban local authority in Leicestershire. It had a high percentage of records that had exact agreement between Valuation Office Agency (VOA) property type and 2011 Census accommodation type.

The most common accommodation type in England and Wales on the 2011 Census was "semi-detached", the second most common was terraced, closely followed by "detached" and then "purpose-built flat". The rarer accommodation types were "flat within a converted or shared house", "flat above/within a commercial building" and a "caravan or mobile structure". The distribution of census accommodation types in Oadby and Wigston is broadly similar to England and Wales as a whole with the two most common types being the same. These two accommodation types ("semi-detached" and "detached") also have the highest linkage rates and one of the highest agreement rates in the linked data for England and Wales as a whole.

One clear difference is that Oadby and Wigston has a lower proportion of flats compared with the linked data for England and Wales. Flats had one of the lower linkage and agreement rates. It may be that the types of properties most prevalent in a local authority influences the agreement rate.

**Figure 3: Agreement between census accommodation type and VOA property type in Oadby and Wigston**



Figure 3: Agreement between census accommodation type and VOA property type in Oadby and Wigston
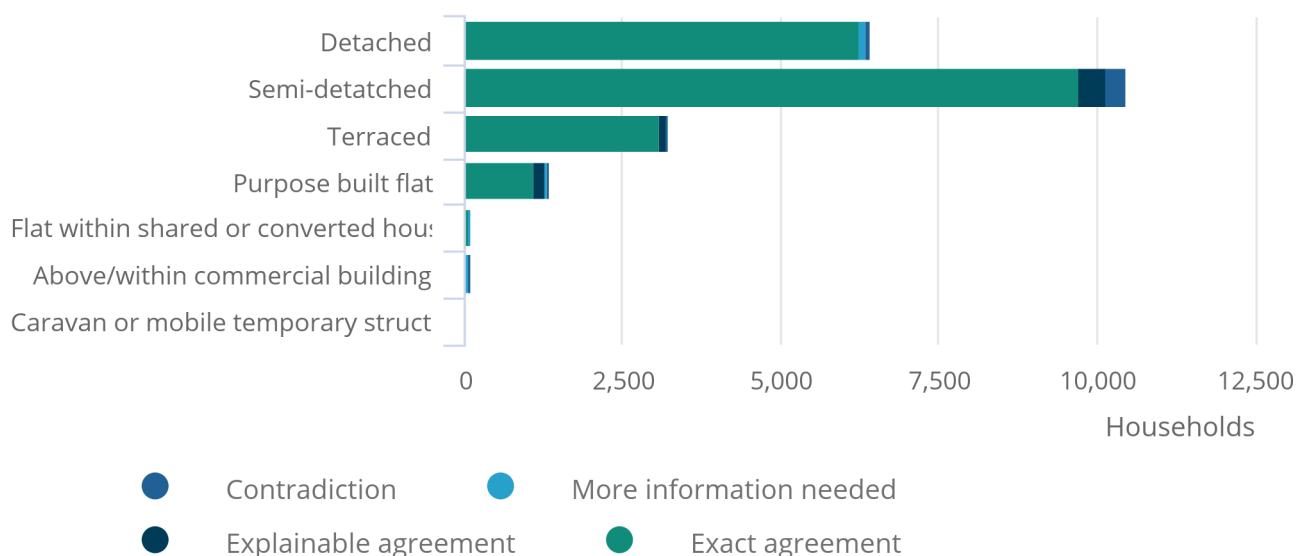
Figure 3 shows the proportion of each agreement category for each 2011 Census accommodation type. All the property types have a similar proportion of each agreement category, with semi-detached having slightly higher explainable agreement and no agreement (contradiction).

A higher agreement means that there is less difference between what the 2011 Census accommodation type reports and what the VOA property type variables record. Therefore, a higher agreement would make transitioning to this data source easier as there would be less impact on users.

## Case study for Newham

Newham is a London borough characterised by grid-plan streets, largely with older terraced housing plus flats. Newham has relatively high demands for private rented and social housing. Newham has the lowest exact agreement (69%) and the highest proportion of contradictions (16%). We are using this as an example of how having lower agreement would look within a local authority.

The main difference between the distribution of accommodation types in Newham and England and Wales overall is a higher proportion of purpose-built flats in Newham. Purpose-built flats have a low exact agreement in the overall data for England and Wales. So, a higher proportion of this accommodation type in Newham is likely to contribute towards an overall low agreement rate.

Similarly, the low level of detached and semi-detached properties (types that usually have high agreement in the rest of the linked dataset) may also contribute to the low agreement rate in Newham, especially as most of the Newham records in these categories have no agreement (contradiction).

**Figure 4: Agreement between census accommodation type and VOA property type in Newham**



Figure 4: Agreement between census accommodation type and VOA property type in Newham
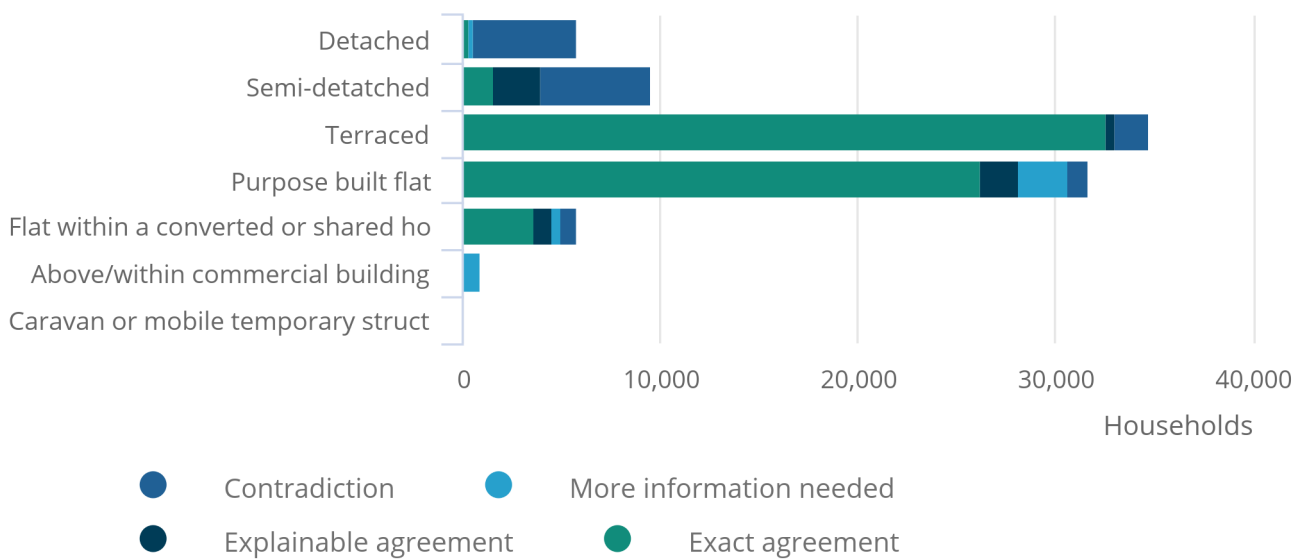
**Source: Valuation Office Agency, Office for National Statistics © Crown Copyright 2015. This work uses research datasets, which may not exactly reproduce National Statistics aggregates**

A lower agreement means there is a larger difference between the 2011 Census accommodation type variable and the VOA property type variable (Figure 4). This means that you would see a larger change in the data if you changed from using the 2011 Census accommodation variable to the VOA property type variable in areas with low agreement. Only a small proportion of local authorities would be affected as 89% of local authorities had exact agreement of over 80%.

# 10 . Next steps

This research demonstrates that there are benefits to using the Valuation Office Agency (VOA) property type. The VOA data are of high quality with high rates of linkage and agreement to the 2011 Census. A main benefit of VOA property type data is that they provide additional breakdowns of flats and allow the separation of bungalows from houses.

In future we may be able to produce census-like analyses by VOA property type, if users would benefit from the breakdowns of property type available from these data. Further research will be required to establish if we can identify properties "within or above a commercial building" from administrative records as is possible with census accommodation type. If possible, this would allow us to have a similar category to the census but from administrative data.

We need to work alongside users of statistics to ensure these data meet their needs and to see how we can address continuity issues where administrative data do not agree with survey-collected information. We plan to publish more of our research on housing statistics using administrative data in the future based on feedback from users.

# 11 . Feedback

We would like to get feedback on these Research Outputs and the methodology used to produce them, including how they might be improved and potential uses of the data. Please email your feedback to admin.based. characteristics@ons.gov.uk. Please include "Housing" in the subject line of your response. We are very interested in understanding what property type categories are likely to be of interest in the future to inform policies, target schemes and monitor changes over time, to ensure we meet user needs where possible. Please let us know:

- What housing information is useful to monitor supply and demand and what level of detail would be required for this?

- What are housing data used for and what specific analyses by property type are of interest?

- How important is it to know if properties are within a commercial building?

Please provide as much detail as you can and email your response to admin.based.characteristics@ons.gov.uk; information provided will be considered in future research.

# 12 . Annex 1: Method used to link VOA data and 2011 Census records

The census dataset we chose to use does not contain imputed values. Imputed values would not be expected to always be true or correct at individual household level but are representative of property characteristics at an aggregate level. Therefore, imputed values are not used in our analysis.

To link the data, addresses were assigned a unique property reference number (UPRN) on both 2011 Census and Valuation Office Agency (VOA) datasets. The same methodology was used in a previous publication comparing number of rooms data between VOA and 2011 Census. We used the same methodology to assign these standard identifiers for both datasets.

Census and VOA records that could not be assigned a UPRN were removed. Records that had a non-unique UPRN (duplicates) were also removed prior to linkage. In addition, records in the VOA property data that were built in or after 2012 were removed in order to match up to the 2011 Census better. Of 23.7 million unique 2011 Census addresses, 22.6 million were able to be linked to VOA property data (see Figure 5).

Some records on both datasets could not be assigned a UPRN, which means we cannot link those records and do not have the information on property type for those records. There were 2.6% of VOA records and 1.7% of census records that could not have a UPRN assigned. Those records were removed before linkage. Note that if one type of property has more records with missing UPRN than others, then that accommodation type will have a lower linkage rate. That can explain why some types of property have different rates of linkage than others.

Some UPRNs appear more than once in the data. In the 2011 Census, this could be because there was more than one household at an address. Alternatively, it may relate to the precision of address information. On the VOA, 246,100 duplicate records and 242,400 from the census were removed before linkage to ensure a one-to-one match.

As the 2011 Census is earlier than the VOA data we have from 2016, we removed any records that were built in or after 2012. Note some existing properties may have been updated in VOA property data since 2011, which may account for some of the difference in agreement that this research finds.

After cleaning was complete there were 22,634,100 total linked records and we used those records to calculate agreement between the 2011 Census accommodation type and the VOA property type.

**Figure 5: Process of cleaning and linking the VOA and the 2011 Census**
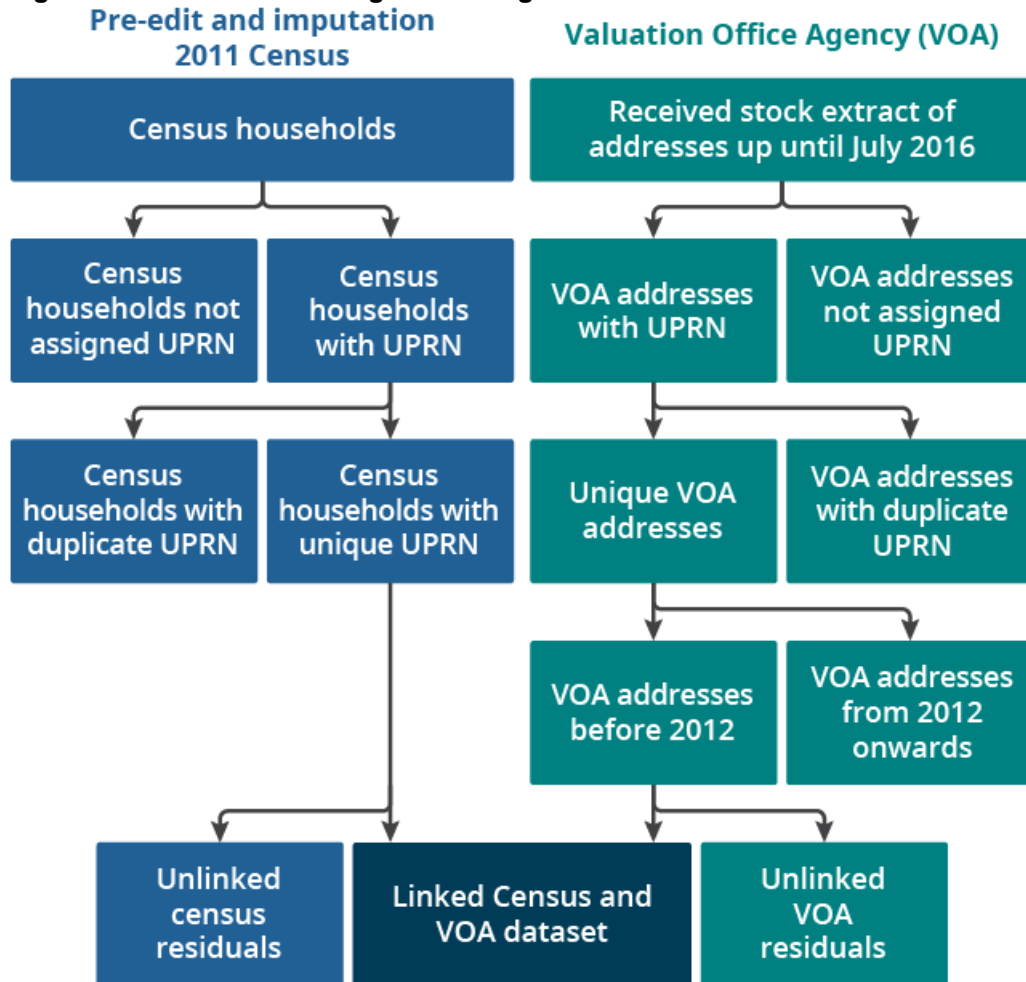


Source: Office for National Statistics

# 13 . Annex 2: Mapping VOA property type onto 2011 Census accommodation type

The census has seven different accommodation types while the Valuation Office Agency (VOA) data records 30 different property types. In this annex we list the seven census accommodation types, their definition and which VOA property types can be mapped onto them.

## Detached

"A whole house or bungalow that is detached" in the 2011 Census dataset is captured by the following categories in the VOA dataset:

- detached house

- detached bungalow

## Semi-detached

"A whole house or bungalow that is semi-detached" in the 2011 Census is captured by the following categories in the VOA dataset:

- semi-detached house

- semi-detached bungalow

## Terraced

"A whole house or bungalow that is terraced" in the 2011 Census is captured by the following categories in the VOA dataset:

- mid-terraced house

- end-terraced house

- mid-terraced bungalow

- end-terraced bungalow

## Flat, maisonette or apartment

"A flat, maisonette or apartment" is separated into two further subcategories in the 2011 Census dataset. These are:

- a flat, maisonette or apartment that is in a purpose-built block of flats or tenement

- a flat, maisonette or apartment that is part of a converted or shared house including bedsits

These categories do not have direct agreement (they do not fully match to the VOA dataset) and are explained in more detail as follows.

## Purpose-built block of flats or tenement

The "flat, maisonette or apartment that is in a purpose-built block of flats or tenement" category from the 2011 Census dataset is captured by the following categories in the VOA dataset:

- purpose-built flat with lift

- purpose-built flat without lift

- purpose-built maisonette with lift

- purpose-built maisonette without lift

There are also four other VOA dataset categories that are a close match but do not directly map across. These are:

- flat in two storey block detached

- flat in two storey block semi-detached

- flat in two storey block mid-terraced

- flat in two storey block end of terrace

## Converted or shared house including bedsits

The "flat, maisonette or apartment that is part of a converted or shared house including bedsits" category from the 2011 Census dataset is captured by the following categories in the VOA dataset:

- flat with lift in group 61[1]

- flat without lift in group 61

- maisonette with lift in group 61

- maisonette without lift in group 61

- non-purpose-built flat, self-contained

- non-purpose-built flat, not self-contained

- non-purpose-built maisonette, self-contained

- non-purpose-built maisonette, not self-contained

2011 Census "accommodation type" of "flat, maisonette or apartment that is in a commercial building" is not shown as there is not a logical equivalent in VOA data.

## Caravan or mobile temporary structure

The "caravan or mobile temporary structure" category from the 2011 Census dataset is captured by "caravan" and "mobile home or house boat" in the VOA dataset.

**Notes for: Annex 2: Mapping VOA property type onto 2011 Census accommodation type**

1. Group 61 was used for flats in converted former commercial and industrial premises. It is a category no longer used by the VOA.

# 14 . Annex 3: List of VOA property types and 2011 Census accommodation types

## The 2011 Census (VOA) has seven different accommodation types. They are:

- A whole house or bungalow that is semi-detached

- A whole house or bungalow that is terraced (including end-terrace)

- A flat, maisonette or apartment that is in a purpose-built block of flats or tenement

- A flat, maisonette or apartment that is part of a converted or shared house (including bedsits)

- A flat, maisonette or apartment that is in a commercial building (for example, in an office building, hotel, or over a shop)

- A caravan or other mobile or temporary structure

## The Valuation Office Agency (VOA) has 29 different property types. They are:

- Detached house

- Semi-detached house

- Mid-terraced house

- End-terraced house

- House in a cluster

- Detached bungalow

- Semi-detached bungalow

- Mid-terraced bungalow

- End-terraced bungalow

- Bungalow in a cluster

- Annexe

- Flat in two storey block detached

- Flat in two storey block semi-detached

- Flat in two storey block mid-terraced

- Flat in two storey block end-terraced

- Purpose-built flat with lift

- Purpose-built flat without lift

- Non-purpose-built flat self-contained

- Non-purpose-built flat not self-contained

- Flat with lift in group 61 only

- Flat without lift in group 61 only

- Purpose-built maisonette with lift

- Purpose-built maisonette without lift

- Non-purpose-built maisonette self-contained

- Non-purpose-built maisonette not self-contained

- Maisonette with lift in group 61 only

- Maisonette without lift in group 61 only

- Caravan

- Mobile home or house boat