

# Evaluation of addressing quality: Census 2021

Summary of the work conducted to build the Census 2021 address frame and an evaluation of the frame's quality.

Contact:  
Census customer services  
census.customerservices@ons.  
gov.uk  
+44 1392 444972

Release date:  
9 January 2023

Next release:  
To be announced

## Table of contents

1. [Main points](#)
2. [Overview of the address frame](#)
3. [Specification for address frame](#)
4. [Creation of Census 2021 address frame](#)
5. [Quality assurance of the Census 2021 address register](#)
6. [Address delta](#)
7. [Total number of addresses](#)
8. [Under-coverage in the Census 2021 address frame](#)
9. [Over-coverage in the Census 2021 address frame](#)
10. [Misclassification in Census 2021 address frame](#)
11. [Duplication in Census 2021 address frame](#)
12. [Coverage of communal establishments with over 50 bedspaces](#)
13. [Related links](#)
14. [Cite this methodology](#)

# 1 . Main points

- The creation of an effective address frame was essential for the Census 2021 collection operation and in the production of high-quality statistics.
- AddressBase Premium (ABP) was the primary source for the frame, supplemented by other administrative data sources.
- The first version of the address frame contained over 27 million addresses across households, communal establishments (CEs) and Special Population Groups (SPGs).
- Over 100,000 additional addresses were added to the address frame by the address delta.
- The statistical and operational design for Census 2021 allowed for continual updating of the address frame during the collection, through information received from the public, our census field officers, or other data sources.
- The address frame met its main objective of providing a comprehensive list of addresses through which to contact the public to complete their census form.

## 2 . Overview of the address frame

All figures in this publication are rounded to the nearest 100 or where below 100 to the nearest 10.

We stated in the [2018 White Paper Help Shape Our Future: The 2021 Census of Population and Housing in England and Wales \(PDF, 967KB\)](#) that we would attempt "to collect information on all usual residents and all persons present at an address on the census date". To achieve this aim, we created a comprehensive address frame for all residential addresses within England and Wales.

We created specific quality targets for the address frame to ensure that a high-quality and cost-effective collection operation could be conducted. These were detailed in our [Design for Census 2021 article](#) and include:

- identifying 99.25% of residential addresses (no more than 0.75% under-coverage)
- achieving no more than 2% under-coverage in any local authority
- identifying 100% of communal establishments with over 50 bed spaces
- achieving less than 1% over-coverage
- wrongly classifying no more than 0.3% of addresses
- including no more than 0.3% duplicate addresses

Although it was important to limit the number of non-residential addresses that were included on the frame, the priority in creating the frame was to limit under-coverage as much as possible. Where there was uncertainty over whether an address was residential or non-residential, the decision was generally made to include these addresses.

Non-residential or derelict addresses could be invalidated and removed from follow-up if the initial contact letter (ICL) was undeliverable, or if we were informed by our census field officers, or the public, that the address was not a residence.

## 3 . Specification for address frame

In 2015, we ran the 2021 Census Topic Consultation to understand our users' requirements for the Census 2021 output and enumeration bases. Read more in our [Census 2021 Assessment of initial user requirements on content for England and Wales: Output and enumeration bases report \(PDF, 661.2KB\)](#).

We decided not to change the household definition to ensure that there was “continuity with the 2011 Census enumeration base”. Census 2021 defined households and communal establishments (CEs) as follows.

A household is: “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.”

A CE is: "An establishment providing managed residential accommodation. 'Managed' in this context means providing full-or part-time supervision of the accommodation." Read more in our [Output and enumeration bases: residential address and population definitions for Census 2021 article](#).

Although the overall definitions for a household and a CE did not change from the 2011 Census, there were slight changes to the types of accommodation included under each category.

The following types of accommodation were changed from being defined as CEs to households in 2021:

- sheltered housing units
- serviced apartments
- hotels, guest houses, bed and breakfasts (B&Bs), inns and pubs with space for fewer than seven guests
- nurses' accommodation

Finally, as part of the address register's design, a third category of accommodation was also created covering special population groups (SPGs). These were residential household accommodations, where tailored engagement needed to be conducted, for example, royal households, embassies and travelling persons. Further information on SPGs can be found in our [Design for Census 2021 article](#).

In the final census data, this type of accommodation is classified as a household. However, for the collection operation, they were enumerated separately and as such existed independently on the address frame.

## 4 . Creation of Census 2021 address frame

We planned to create an address frame by using several different sources of information, such as:

- administrative data sources
- desk-based clerical work
- field address check

Because of the coronavirus (COVID-19) national lockdown, it was not possible to conduct the proposed field address check. Instead, we increased the amount of resource placed into our desk-based clerical checking activities, which involved using online resources such as Google Maps, web searches, and investigations into land registry information.

The increased resourcing of our desk-based clerical work led to our address checking processes becoming far more efficient and effective.

It also provided us with insights about addressing for property types such as houses of multiple occupation, which we may not have obtained from a field address check. We built on what we had learned during this exercise by establishing a similar team to assist in the resolution of addresses during the census collection operation.

## Household address list

The household section of the address frame was developed using AddressBase Premium epoch 77, which was taken as of June 2020. We chose to use AddressBase Premium as the main source for the household section of address frame because it was and still is the most authoritative, current and accurate dataset on addresses in England and Wales.

AddressBase Premium uses data from the Local Land and Property Gazetteers (LLPGS) in conjunction with a range of address intelligence sources, such as from the Valuation Office Agency (VOA), Royal Mail and Ordnance Survey. Read more in our [Census 2021 Address Checking report \(PDF, 251KB\)](#). The data were taken in June 2020 because of the time needed to print initial contact letters. We created address deltas (see Section 6) and received weekly updates on new addresses from Geoplace (supplier of AddressBase Premium) to ensure that we were aware of all addresses that became active up to Census Day.

The address records from AddressBase Premium that made up the initial household address frame were classified into three different extracts, which were called:

- extract one – these were addresses that had a residential classification in AddressBase Premium and a Council Tax link (view the [AddressBase classification codes \(zip, 625KB\)](#))
- extract two – these were addresses that did not have a residential classification in AddressBase Premium but did have a Council Tax record
- extract three – these were addresses that had a residential classification in AddressBase Premium but did not have a Council Tax record

Table 1 provides further detail on the number of records contained within each extract.

Table 1: Number of addresses within each extract of the household address frame for Census 2021

Type of record	Count
Extract one	25,776,300
Extract two	154,200
Extract three	1,122,800

Source: Office for National Statistics – Evaluation of addressing quality: Census 2021

While extract one and extract two addresses were automatically added onto the household address list, further clerical work was conducted to resolve the extract three addresses. The Office for National Statistics' desk-based address resolution team (DART) reviewed these addresses and decided whether they should be added.

## Communal establishment and special population group address list

AddressBase Premium could not be used as the primary dataset for the creation of the communal establishment special population group (CE-SPG) address frame. This was because “there [were] known limitations [with] the AddressBase Premium product [around] the quality of the addresses for communal establishments”. Read more in our [Census 2021 Address Checking report \(PDF, 251 KB\)](#).

Therefore, a variety of administrative data sources were used to assist in the building of the Census 2021 CE-SPG address frame. Read more in our [Administrative Data Used in 2021 Census report \(PDF, 806KB\)](#).

While most communal establishments were enumerated at an establishment level, this was not the case for halls of residences. Obtaining and using individual room addresses was considered better for the collection operation, as it would allow for targeted follow-up on non-response. Read more in [the UK Statistics Authority's Design of Address Frame, Collection and Coverage Assessment and Adjustment of Communal Establishments in 2021 Census paper](#).

To facilitate this, we asked local authorities (LAs) to provide room-level information for student halls, where they were able to do so by updating their information in AddressBase Premium. The Communal Establishment Address Resolution Team (CEART), a dedicated team established to conduct this work, also contacted universities and private halls of residence providers to obtain room-level information, where it did not exist in either AddressBase or administrative data supplied by Cushman and Wakefield. In total, more than 400,000 room-level addresses were obtained through the CEART Operation.

The CEART team also identified the classification type and capacity of other types of CEs. This was to reduce the likelihood of misclassification and to ensure the correct number of forms were delivered to each establishment. The results of this clerical work were as follows:

- 2,800 records were reviewed by CEART
- 1,900 records were deemed to be classified as the correct establishment
- 500 records changed classification to another establishment type or to a household
- 400 records were removed from the frame

## 5 . Quality assurance of the Census 2021 address register

A variety of quality assurance measures were put in place to ensure the accuracy of the address frame. We made a comparison with independent sources, such as Personal Demographic Service (PDS) data to identify missing addresses. We also reviewed Communal Establishment Address Resolution Team (CEART) collected data to identify duplicate room-level addresses. Read more in our Census 2021 [Address Checking report \(PDF, 251KB\)](#).

The results of the comparator analysis provided us with reassurance that the address frame had not missed a large cluster of addresses.

## 6 . Address delta

In a similar manner to the 2011 Census, an updated address file called the “address delta” was delivered to enhance the quality of the Census 2021 address frame. This was supplied for three reasons, which were:

- to add any new build addresses
- to remove any inactive addresses
- to correct any misclassified addresses

Epoch 80 address data from AddressBase Premium was used for the address delta. The address delta file was delivered in January 2021 and the following changes were made to the address frame:

- 103,800 new addresses were added to the household address list
- 2,200 new addresses were added to the communal establishment special population group (CE-SPG) address register
- 65,700 addresses were removed because they were now considered invalid (primarily as a result of becoming “historical” addresses)

## 7 . Total number of addresses

The final version of the Census 2021 address frame contained the following number of addresses detailed in Table 2.

Table 2: Number of establishments in the address frame by address type for Census 2021

Address type	Count
Households	27,053,200
Communal Establishments and Special Population Groups (CE – SPG)	49,900

Source: Office for National Statistics – Evaluation of addressing quality: Census 2021

All household addresses included in the final version of the address frame had printed contact materials that were dispatched prior to Census Day, including those added in the address delta file. Given that this address list was finalised in advance of March 2021, and therefore might potentially miss new addresses added after Epoch 80, we collaborated with GeoPlace to continue to update our address lists. Although these addresses did not receive initial contact letters, there were followed up in subsequent waves of contact. Read more in our [Design for Census 2021 article](#).

## 8 . Under-coverage in the Census 2021 address frame

We were able to assess the coverage of the address frame for household and special population group (SPG) addresses by comparing them with the Census Intelligence Datasource (CID). This is an address-level dataset, which combines several census and administrative data sources together to provide a wide range of insights on addresses in England and Wales, and therefore provides a comparative estimate for the number of households.

For communal establishments (CEs) in the address frame, it was not possible to provide a similar assessment for the coverage because of the lack of a reliable independent comparator data source.

The CID was developed to assist in the quality assurance of census results. The administrative data sources used in the CID are as follows:

- Housing Association Vacant Property data
- Valuation Office Agency (VOA) data
- Personal Demographic Service (PDS) data
- Council Tax data
- Higher Education Statistics Agency (HESA) data
- English School Census (ESC) data
- Electralink data
- Xoserve data
- Electricity Central Online Enquiry Service (ECOES) data

The original address register for households and SPGs was compared with the CID to provide an estimate on the level of under-coverage. The average under-coverage of addresses was:

- 0.36% for England and Wales
- 0.35% for England
- 0.49% for Wales

These figures show that while some under-coverage was present, this was well within the quality target set for Census 2021. As such, the original address frame met its quality target of limiting under-coverage across England and Wales to less than 0.75%. Wales did have a slightly higher rate of under-coverage than England, and further investigative work may be needed to understand the reasoning behind this.

## Local authorities

Comparator analysis was conducted between the CID and the household SPG address frame. This was to determine whether under-coverage was below 2% for each individual local authority (LA).

For England, 99.7% of LAs met the quality target of no more than 2% under-coverage, and 100% of LAs in Wales met the quality target of no more than 2% under-coverage. The only LA that did not meet the target was small in size, and the number of addresses missed was fewer than 50.

Addresses not included on the frame were subsequently added during the Census 2021 collection operation. More clerical work may be considered to improve coverage in future for this particular LA.

## 9 . Over-coverage in the Census 2021 address frame

To ascertain the level of over-coverage within the original address frame, we examined the following invalidation reasons provided by the census field officers:

- duplicate
- non-residential
- derelict-demolished
- does not exist
- cannot find property
- merged
- demolished
- under construction
- derelict
- undelivered as addressed
- un-addressable

These codes were chosen because:

- nobody could have been currently living in these properties
- the address did not exist
- the address was a duplicate

The investigation into over-coverage provided the following results:

- average over-coverage for England and Wales – 2.1%
- average over-coverage for England – 2.1%
- average over-coverage for Wales – 1.7%

The findings show that the address frame did not meet its quality target of limiting over-coverage to less than 1%, with both England and Wales encountering over-coverage higher than the set target. More analysis was conducted to understand the effect that over-coverage had on each address type, there were:



- 50 local authorities that had over-coverage of household and special population group (SPG) addresses between 0% and 1%
- 130 local authorities that had over-coverage of household and SPG addresses between 1.01% and 2.00%
- 100 local authorities that had over-coverage of household and SPG addresses between 2.01% and 3.00%
- 60 local authorities that had over-coverage of household and SPG addresses higher than 3.01%

Most local authorities had an over-coverage of more than 1% for household and SPG addresses. The most common field officer invalidation reason for household and SPG addresses was “under construction”, with 61% of over-coverage addresses having this outcome code.

The primary cause for the level of over-coverage witnessed for households and SPGs was the inclusion of extract three addresses in the address frame (see Section 4 for explanation of extracts). Of the over-coverage household and SPG addresses, 78% were from this extract. As discussed, the decision was made to include addresses on the address frame if they had the potential to be residential and occupied by the start of census collection in March 2021. Given the high proportion of over-coverage in addresses still “under construction”, it may be that this number was inflated by the ongoing coronavirus (COVID-19) pandemic delaying the completion of building projects. More research is needed to understand this.

## 10 . Misclassification in Census 2021 address frame

For Census 2021, addresses were either classified as households, special population groups (SPGs) or communal establishments (CEs). An addresses classification was important as it determined the type of questionnaire it was sent.

When referring to misclassified addresses, we examined household addresses that were misclassified as CEs, on the original frame, and CEs misclassified as households. As SPGs are classified as households in the final census data, we did not look at misclassification of these addresses.

As CEs cater for larger populations, sub-units of a CE could be misclassified as multiple households, or the whole CE could have been wrongly classified as a single household. Consequently, addresses could be both misclassified as well as being a duplicate of another address (see Section 11).

The processes put in place to correct the classification of an address were that census field officers could change an addresses classification during follow-up visits – if a field officer identified a case had been misclassified, they invalidated the case and created a new one with the correct address type. Furthermore, clerical work was conducted after the creation of the address frame, and prior to the processing of census data, to resolve instances of misclassification.

The results of the investigation into misclassification were that, in total, 43,100 addresses were misclassified. The vast majority of these were sub-units of CEs, for example, flats or rooms within halls of residence, that were incorrectly included as households in the original address frame.

This figure equates to 0.16% of all addresses. The quality target for misclassified addresses was less than 0.3%, so the address frame met its quality target.

## 11 . Duplication in Census 2021 address frame

Every residential property on the address frame had a unique identifier so that non-response follow-up visits could be targeted with individual properties. As such, all addresses were given a Unique Property Reference Number (UPRN), which was either provided by AddressBase Premium or by the Office for National Statistics (ONS).

Although all UPRNs are unique, duplicate entries could be included in the following ways:

- an address might be included on the frame for the flats and the “shell” address for a property
- an address might erroneously appear more than once in AddressBase Premium, with different UPRNs
- both an updated and “historical” (no longer in use) version of an address might be included under different UPRNs
- sub-units (rooms, flats, blocks) of communal establishments (CEs) might be included in addition to the overall establishment (see Section 10)

We used several sources of information to identify the level of duplication in the original frame. Firstly, we used information provided by census field officers who identified duplicate addresses in the census collection operation. Furthermore, clerical work, conducted prior to the processing of census data, also identified instances of duplicate address.

Field officer invalidation identified that there were 78,600 duplicate addresses in the frame, while clerical work identified a further 36,400 duplicate addresses.

The analysis revealed that 0.42% of all addresses on the original address frame were duplicates. As such, the address frame did not quite meet its quality target of including no more than 0.3% duplicate addresses.

The primary cause of duplication was the existence of multiple entries for the same address in AddressBase premium, which had completely different UPRNs. More work will be conducted in the future to reduce the instances of duplication in the address frame.

## 12 . Coverage of communal establishments with over 50 bedspaces

This quality target was not able to be assessed because the census does not collect bed space information for each communal establishment (CE) as highlighted in our article, [Estimation and Adjustment for large communal establishments \(CEs\)](#). This target was set, in principle, to ensure that all large CEs were captured on the address frame. Research was conducted to identify whether the original address frame contained all large CEs. Read more in our [Communal establishment \(CE\) estimation and adjustment: Census 2021 article](#).

Of the roughly 5,000 addresses that were factored into large CE estimation, approximately 470 appeared to not exist on the original CE address frame. Further clerical work was completed to ascertain whether these addresses were missing from the address frame or were being captured under a different address entry. The results of this clerical work were as follows:

- roughly 52% of the addresses were on the CE frame under a different unique property reference number (UPRN) to that which existed in the final census data
- roughly 35% were incorrectly classified as households on the original address frame and so existed on the household (HH) address register
- roughly 13% were completely missing from the address frame

While some large CE addresses were missed from the original frame, they were either identified during the collection operation or in estimation. Work was conducted to ensure that the correct number of residents were recorded at each establishment.

Care homes was the most common type of large CE that was missing completely from the address frame. The primary reason for this was determined to be quality issues with the unique property reference number (UPRN) supplied in the administrative data used for the building of the address frame. This led to these addresses not being able to have geographical features added to them and consequently led to their exclusion.

Roughly 27% of the missing establishments had either a non-residential AddressBase Premium classification code or a historical UPRN. These addresses might not have been included because they did not have an AddressBase classification code, which would have helped identify them as a CE.

Several of the missing halls of residences had been identified in the Communal Establishment Address Resolution Team (CEART) operation. However, a valid UPRN was unable to be identified for these buildings, which led to their exclusion from the original address frame.

## 13 . Related links

[Design for Census 2021](#) Web page | Released 2021 How we have planned and designed Census 2021 from start to finish, to ensure results are of high quality and fit for purpose.

[2021 Census Address Checking \(PDF, 252KB\)](#) Report | Released 2021 Summation of clerical work conducted to create the address frame.

[Administrative data used in Census 2021, England and Wales](#) Methodology | Released 15 September 2022 Summary of the administrative data used in Census 2021.

[The 2021 Census Assessment of initial user requirements on content for England and Wales: Output and enumerations bases report \(PDF, 646KB\)](#)  
Report | Released May 2016  
Summary of the opinions provided by bodies on the proposed Census 2021 output and enumeration bases.

[Communal establishment \(CE\) estimation and adjustment: Census 2021](#)  
Methodology | Released 5 January 2023  
Methodology and results of the small and large communal establishment (CE) estimation processes for Census 2021.

[Design of Address Frame, Collection and Coverage Assessment and Adjustment of Communal Establishments in 2021 Census](#) Report | Released August 2021 Summary of the Census processes for enumerating communal establishments.

## 14 . Cite this methodology

Office for National Statistics (ONS) released 9 January 2023, ONS website, methodology, [Evaluation of Addressing Quality: Census 2021](#)