2021 Census Count Challenges Hack Day Report 24th February 2020

Office for National Statistics

Tanita barnett / Jon Wroth-smith

2020

Contents

[Summary 2](#_Toc34069287)

[Scope of the day 3](#_Toc34069288)

[Group sessions 3](#_Toc34069289)

[Attendees 4](#_Toc34069290)

[Key findings from each area 5](#_Toc34069291)

[Group session 1 5](#_Toc34069292)

[Group session 2 9](#_Toc34069293)

[Recommendations and next steps 13](#_Toc34069295)

# Summary

The 2021 Census will make use of a broader range of data sources than any previous census. This includes administrative, survey and commercial data throughout the design from the initial construction of an address frame through to the quality assurance and validation of final estimates.

Following the collection rehearsal in late 2019, we will be finalising the statistical design of the 2021 Census to optimise response and to ensure that resulting census estimates are fit for purpose when published as National Statistics.

This paper provides an overview of a hack day held in February to consider whether any other data could be used as part of the design specifically within:

1. The construction of the address frame
2. The collection operation
3. The estimation and quality assurance of final estimates.

The outcome is a list of potential sources which will be prioritised, and an assessment made of whether they can be accessed ahead of the design being finalised.

A hack day approach was used as a means of conducting initial research into potential data sources using experience from across the office. The approach has been successfully used in finding innovative data sources in the transformation of ONS statistics covering topics as diverse as meat prices, obesity and migration.

Attendees at the hack day were drawn from areas across ONS. This included the 2021 Census Statistical Design and Stakeholder Engagement teams, Methodology, Data Architecture (with expertise in web scraping), Data as a Service (responsible for the acquisition of data), Population Statistics Division, and Statistical Research and Design (conducting research on the use of non-survey data to transform the social statistics system).

# Scope of the day

The aim of the hack day was to consider whether additional data sources not currently part of the design had value in improving the address frame, collection operations and in estimation/quality assurance. While the hack day did consider all groups, focus was given to some of the population groups which have already been identified as Target Action Groups (TAG). In the census design TAGs have been as those groups for whom the standard design needed to be adapted in order to achieve the quality standards required for census outputs.

Not all TAGs were considered in the hack day (for example illegal migrants and rough sleepers). Other groups were added given experience from the 2019 rehearsal and development of the design overall. The groups considered on the day were:

1. Beds in sheds
2. The Gypsy and Traveller community
3. Second homes
4. Gated communities
5. Sofa surfers
6. Multigenerational households
7. Missing individuals in enumerated households
8. Communal Establishment types and completeness
9. Individuals temporarily staying in accommodation without usual residents

Attendees were asked to develop ideas for data sources that could be implemented into the 2021 Census design. It was recognised that there are inevitably limitations to whether complex, personally identifiable sources at record level could be fully researched, acquired, processed and integrated into the design (in addition to cost implications). The hack day was not constrained by these considerations, noting that a range of acquisition options were available including record level, low level aggregates, and web scraping.

Definitions of each of the groups above were given to the attendees to help aid research into new sources and all had access to the Target Action Group information. The ONS Information Asset Register was also available as a reference to the data already acquired by ONS for other statistical purposes and which could be reused within the census design. Groups conducted on-line research during the hack day to identify potential sources and to consider published research based on the sources of interest.

# Group sessions

The attendees were split into 5 groups for the first group session which covered:

1. Beds in sheds
2. The Gypsy and Traveller community
3. Second homes
4. Gated communities
5. Sofa surfers

Each group then fed back their findings to the wider group to allow for further thoughts/ideas to be captured from all attendees.

The attendees were then split into 4 groups for the second group session held in the afternoon which covered:

1. Multigenerational households
2. Missing individuals in enumerated households
3. Communal Establishment types and completeness
4. Individuals temporarily staying in accommodation without usual residents

As with the first group session, a similar feedback session followed. The groups recorded their ideas for data sources using post-it notes and flipchart paper.

# Attendees

|  |  |
| --- | --- |
| Jonathan Wroth-Smith  Cal Ghee  Jaspreet Gakhal  Harry Wrightson  Tanita Barnett  Yasmin Hoque  Bethany Fitzgibbon  Greg Tilsar  Orlaith Fraser  Alexandra Christenson  Nathan Williams  Pete Large | Census Statistical Design |
| Owen Abbott  Gareth Powell  Abu Hossain  Rhiannon Yapp | Methodology |
| Lara Phelan  Neil Townsend | Stakeholder Engagement |
| Tamsin Ribton-Turner | Data Architecture |
| Steve Smallwood | Population Statistics Division |
| Ali Dent | Statistical Design and Research |
| Geum Young Min | Data as a Service |
| Carla Bushell | People and Business Services |

# Key findings from each area

## Group session 1

|  |  |  |  |
| --- | --- | --- | --- |
| Area | Address Frame | Collection | QA |
| Beds in sheds | * **Number of people per address earning on Pay As You Earn (PAYE)** * Satellite images * **Patient Demographic Service (PDS) and use of NHS/health data sets – any reference to ‘annex’ for a registered address** * Any type of business which might have resident person * Should we send some letters/forms to ‘high risk’ non-residential properties * Complaints to councils (planning info) * Does address framework cover alternative addresses which aren’t accepted as addresses? Could there be a flag for collection? | * Service based approach i.e. job centre, mobile health centre * Importance of field staff training * Dummy forms could use info on where beds in sheds are suspected * **Field staff asked to be particularly aware in areas where annexes prevalent** * Census Coverage survey (CCS) already asked to probe to find such addresses | * Flag from field staff to say bed in shed found * Council data on beds in sheds * Mobile phone data * Bus passes (unlikely to have a big overlap with beds in sheds) * Charities in local areas – community groups * **Electricity/water/gas usage per person linked to census data** |
| Gypsy and Traveller  “not living in household address”  Could be:  Site – Authorised/non-authorised or mobile | * **Use school census to identify addresses of Gypsy and Traveller children, check with address register for sites** * MHCLG frame from local authority (LA) returns unit level data on Gypsy and Traveller/caravan count * **Planning permission regularly rejected** * Specialist insurance * Satellite imaging changes in caravan locations * Address check of new addresses need to be aware some could be Gypsy and Traveller communities | * Any ad hoc surveys by LAs to monitor health, accommodation etc. * Field operation paying gatekeepers * **Scrape social media – report of Gypsy and Traveller communities around census day** * DVLA data – people/owner, addresses | * Department for Education – temporary and permanent exclusion data in schools for Gypsy and Traveller children. * Prisoner data to identify Gypsy and Traveller individuals in prison * Gypsy and Traveller events around census day potential QA resource later * Gypsy-traveller.org help/advice for community * What about Gypsy and Travellers not defining as Gypsy and Travellers something to identify the extent of this? |
| Second homes | * Council tax second residence discount – includes holiday homes * Holiday websites – Airbnb * Holiday lets – NDR * TV licence * **Stamp duty – second homes pay an additional 3%** * Second mortgage | * Completion of dummy forms * Referencing council tax as corroborating field finding * **Asking neighbours whether address is a second residence** * **Identify areas where second residences are particularly common – highlighting the importance of asking neighbours** | * English Housing Survey |
| Holiday home websites: travel supermarket, Airbnb etc  Possible data  **Council tax**  **Second residence discount**  **Named people paying in 2 locations. Non-domestic i.e. rented out as business = holiday home**  Insurance data – multiple households/contents insurance  TV licence – 2 licenses in 2 places for same individual  Mobile phone – same individual spending significant time at two addresses  Mortgage lenders. Second mortgage for those planning to  HMRC stamp duty  Buy to let to rule out second home  MLAR – central mortgage data  Land registry – multiple owners  **Utility usage - utilities used only at weekends, working week**  English Housing Survey for QA  Potential situations discussed  1 – Whole household second home e.g. family home in Berkshire, flat in Westminster for work during the week  2 – Partial household second home e.g. Family home in Hampshire, live in shared house in Kensington & Chelsea during week  3 – Foreign resident second home (UK), live in Paris but have a flat in London. | | |
| Gated communities:  Discussion on types of gated community:  Aged 50-70 ‘downsizing’ retirement homes, ‘prestige developments | * **Rightmove key word** * 2011 Census data * ONS Social Survey * **Companies that manage a gated community** * Commercial estate agents * Home insurance * Refuse collections * Delivery companies (UPS) * **Royal mail aware** * “Good things” data * **Emergency services** * Royal institute of chartered surveyors * English planning authority – planning applications * New developments * Social housing * Leasehold information * Glasgow University – urban studies * Union of concierge * ‘Access to or use of buildings’ MHCLG * Sarah Blandy university of Leeds (2007) * Health providers for older people * **Atkinson (2000) paper Home Office research into Gated Communities** * Local authority private roads * Google street view search * Concierge/security job adverts * **Key words to search in SharePoint – published dissertation** * Other NSIs approaches e.g. USCB * Local authorities' data * Retirement villages websites * Caravan sites * Golf clubs and private residents * Retirement gated communities | * Tranche 1 field staff – report **back findings of gated communities** * Write to owners/companies of gated communities * Info for area managers * Sending letters rather than follow up |  |
| Sofa surfers | * Sofa surfers in non-residential addresses – businesses, covered by beds in sheds? | * Tailor messages from field staff * Localised targeting via social media – most/all will have online access * Youth sofa surfers – DfE data from schools * **Ask academics – Herriot Watt University research on sofa surfers** * Relationship breakdown estimates * **Both Collection/QA** * LA housing waiting lists LAHS section C * Student waiting for accommodation * Shelter/Crisis * **Couchsurfing data – webscrape** * Social services data on children sofa surfers * English Housing Survey - % of households living in another household’s accommodation as an additional family and couldn’t afford to buy or rent own housing * Electoral role data: no fixed address | * Visitor data from census form |

## Group session 2

|  |  |
| --- | --- |
| Multigenerational households | Less likely to be a problem as with:   1. Single person (male household) 2. Unrelated people in household etc.   Use of admin sources i.e. PDS, pensions 65+  Will the adjustment process retain the distribution of household composition e.g. having enough households with older people living with younger generations |
| Missing individual in enumerated households | Census Non-Response Link Study  **Paradata on eQ where people reduce number of people in household**  Special attention to 5 person households on paper, requested continuation not completed  **Can we learn anything from 2011 CCS link or from Longitudinal Study**  Special analysis of under 1s – birth by month  Locating number of rooms – particularly with students/student areas  **Pertinent issue with households of multiple occupancy. HMO register from Local Authorities already published**  Using admin data on household size/patterns (may only be useful in extreme cases)  Can we sample for follow up addresses where either number of rooms is far higher or admin data suggests numbers are higher. |
| CE types and completeness | Students  Students in halls:   * HESA * Planning applications * Fire services * 2011 census   NUS  Student loans  Webscraping bedspaces data from university halls – Data Science Campus algorithm  Care homes  Deaths in care homes  Carehomes.co.uk webscraped bedspaces  Care worker adverts  [www.housingcare.org](http://www.housingcare.org) webscraped bedspaces data  Boarding schools  Department for Education  School Census  Ofsted reports  Children’s homes  Births  Maternity homes  Religious establishments  Protective security funding schemes  Home office scheme for religious establishments  Councils for:   * Church of England * Catholic * Muslim * Sikh * etc.   Webscraping info from dioceses websites  Detention centres  Immigration removal centre – Home Office |
| Individuals temporarily staying in accommodation without usual residents | Boat-sitting  Intersections of business and residential property. Check business rate council tax vs domestic rates  CCS also covers some commercial properties likely to have residents. Plant recapture idea?  Other people not officially guardians, e.g. sleeping at back of shop. May be under the radar anyway, would we/could we find and chase up  What is not going to be dealt with by the CCS but big enough to worry about?  **Property guardians – website adverts**  Recent property conversions. Commercial to residential.  Property guardians messaging to guardian organisations/agencies to cascade out to the guardians  Tree diagram provided below: |

### Individuals temporarily staying in accommodation without usual residents

# Recommendations and next steps

An initial consideration has been made of recommended priorities for further investigation. Given the range of types of data and information considered, initial priorities have been categorised by type.

Next steps for prioritises are summarised as:

* Report Hack Day findings to External Assurance Board
* Evaluate options for data acquisition with relevant teams (Data as a Service, Data Engineering, 2021 Census teams) making recommendations on whether to pursue as direct acquisition or webscraping.
* Evaluate use within existing design to consider where data can be used and whether design changes are required.
* Proposals provided to 2021 Census Design Authority Board

**Recommended initial prioritisation**

**Use of data already acquired or in the process of being acquired**

* Number of people per address earning on Pay As You Earn (PAYE)
* Patient Demographic Service (PDS) and use of NHS/health data sets – any reference to ‘annex’ for a registered address
* Use school census to identify addresses of Gypsy and Traveller students, check with address register for sites
* Council tax: Second residence discount, named people paying in 2 locations. Non-domestic i.e. rented out as business = holiday home

**Use of data within 2021 Census**

* Field staff asked to be particularly aware in areas where annexes prevalent
* Field staff asking neighbours whether address is a second residence
* Identify areas where second residences are particularly common – highlighting the importance of asking neighbours
* Tranche 1 field staff – report back findings of gated communities
* Can we learn anything from 2011 CCS link or from Longitudinal Study?

**Potential new sources**

* Electricity/water/gas usage per person linked to census data and utility usage - utilities used only at weekends, working week
* Stamp duty – second homes pay an additional 3%
* Pertinent issue with households of multiple occupancy. HMO register from Local Authorities already published
* Companies that manage a gated community
* Royal mail aware of gated communities
* Emergency services access to gated communities
* Paradata on eQ where people reduce number of people in household

**Potential new sources (scraped)**

* Planning permission regularly rejected
* Scrape social media – report of Gypsy and Traveller communities around census day
* Rightmove key word
* Couchsurfing data – webscrape
* Property guardians – website adverts

**Existing research**

* Atkinson (2000) paper Home Office research into Gated Communities
* Key words to search in SharePoint – published dissertation
* Ask academics – Herriot Watt University research on sofa surfers