Economic Commission for Europe
Conference of European Statisticians
Group of Experts on Population and Housing Censuses
Twentieth Meeting
Geneva, 26–28 September 2018
Item 2 of the provisional agenda
Methodology, new data sources including big data

Approach to using alternative data sources to support the 2021 Census in England and Wales

Note by the 2021 Census Statistical Design, Office for National Statistics (United Kingdom of Great Britain and Northern Ireland)*

Summary

For our 2021 Census we are looking at how we plan to integrate alternative data sources end-to-end in order to optimise the quality of the census and better meet user needs. Plans are at varying degrees of maturity regarding use in preparation, collection, processing and outputs. This document sets out our approach and aims, design principles and criteria for inclusion, and an update on where we are with each aspect. It also presents results of tests where available, including potential uses that have been ruled out.

* Prepared by Cal Ghee, Garnett Compton and Marie Haythornthwaite.
I. Introduction

1. Administrative and survey data from local authorities and central government played a key role in the 2011 Census design. These data were used extensively in the quality assurance stage of the 2011 Census estimates and in the planning of the operation. The Office for National Statistics (ONS) intends to use a wider range of administrative, survey and other alternative data to support the research, design and operation of the 2021 Census for England and Wales, as well as producing a wider range of outputs (than just core census based statistics) by integrating census and administrative data in line with the National Statistician’s recommendation in March 2014 on the future of the Census and Population Statistics (ONS, 2014).

2. This document sets out our initial thinking on how these data could be used as part of the design for the 2021 Census in England and Wales. It’s important to note that this thinking remains indicative at present. The exact ways in which we’ll use these data will depend on:

   (a) The data available;

   (b) The outcome of our research into the quality of available data;

   (c) The development of methods to harness the power of the data.

3. This document doesn’t comment on the wider use of administrative data in ONS to improve the quality of population, migration, economic and social statistics or the potential to move to an admin data led census (ONS, 2018) in the future in support of the ONS’ Better Statistics, Better Decisions strategy (UK Statistics Authority, 2014). The National Statistician will make a recommendation about the future of the Census after 2021 in 2023.

4. The UK Census is made up of 3 separate operations: Scotland and Northern Ireland are developing their own design based on the most appropriate methods. ONS, National Records Scotland (NRS), Northern Ireland Statistics Research Agency (NISRA) and Welsh Government (WG) ensure the methods and outputs are harmonised to provide coherent UK statistics. This document only covers the work of ONS on the census for England and Wales.

II. Objectives of the 2021 Census

5. Administrative, survey and other alternative data sources have the potential to help meet 2 specific objectives of the 2021 Census, namely:

   (a) To produce statistics of the right quality and timeliness to meet user needs;

   (b) To produce integrated outputs from census, administrative and survey data.

6. For example, administrative, survey and other alternative data could be used to support census planning and preparation, collection and processing operations. These include validation and quality assurance as well as the creation of outputs and quality reporting. Further details are set out in Chapter IV of this document.

7. Using alternative data sources, methods can be developed and implemented to produce new and enhanced 2021 Census outputs (for example, linking administrative data on income to census outputs to produce small area, multivariate income statistics). Further details are included in chapter IV section C sub-section 1 below.
III. What are alternative data: administrative, survey and other sources

8. ONS is considering a range of different additional data sources to help meet the objectives of the 2021 Census. These include the following.

A. Administrative data

9. This refers to information collected primarily for administrative reasons (meaning not initially for statistics or research). This type of data is collected by Government departments and other organisations for uses such as registration, transactions and record-keeping usually as a by-product of delivering a service. Administrative data are often used for operational purposes and their statistical use is usually secondary.

B. Survey data

10. These data are gathered from statistical surveys including earlier censuses, and surveys such as the Labour Force Survey (ONS, 2015) and Annual Population Survey (ONS, 2012a).

C. Other sources

11. In the context of this document, we’re using this as a collective term referring to all data not traditionally used within ONS. These include:

   (a) **Big data:** ONS defines big data as large, often unstructured data sets that are available, potentially in real time, which are difficult to process efficiently using traditional methods and technologies. The amount and variety of data available is growing rapidly. There’s a wide range of data available in many formats, including audio, video, computer logs, purchase transactions, sensors and social networking sites. Some of these data are freely available on the web, whereas others are held by the private sector.

   (b) **Management information:** This describes aggregate information collated and used in the normal course of business to inform operational delivery, policy development or the management of organisational performance. It’s usually based on administrative data but can also be a product of survey data. The terms administrative data and management information are sometimes used interchangeably.

   (c) **Metadata:** These are data that describe or define other data. It means anything that users need to know to make proper and correct use of the real data, in terms of reading, processing, interpreting, analysing and presenting the information. Therefore, metadata includes file descriptions, codebooks, processing details, sample designs, fieldwork reports, conceptual motivations etc.

   (d) **Paradata:** These are data that describe the process by which the data were collected, normally through a survey. Examples of paradata topics include:

      (i) The times of day responses were submitted;

      (ii) Time taken to complete the questionnaire;

      (iii) Number of attempts to complete the questionnaire;
(iv) How many times field officers called at non-responding addresses, day of week and time of day, how many times they made contact and whether a response was subsequently received;

(v) Mode of communication (such as phone, web, email, or in person).

Therefore, there are paradata about each observation in the survey. These attributes affect the costs and management of a survey, the findings of a survey, evaluations of field officers and inferences someone might make about non-respondents.

12. The following chapters discuss some of the main areas we are looking at in our 2021 design, but do not include all the confirmed or potential uses.

IV. **How alternative data can help meet 2021 Census objectives**

13. Running a census can be simplified into 3 main stages as shown in Figure I. The following chapters give more detail on some of the potential uses of alternative data in these stages. We are still in the research phase of the census programme, so not all of these possibilities will make it into the final design for 2021.

Figure I  
**Potential uses of alternative data in the 3 stages of the end-to-end 2021 Census operation**

<table>
<thead>
<tr>
<th>Time</th>
<th>INTEGRATION OPERATION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PREPARE &amp; COLLECT</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Helping to create address frame in advance***</td>
</tr>
<tr>
<td></td>
<td>Hard to count; field workload allocation***</td>
</tr>
<tr>
<td></td>
<td>Collection Target Pops; predicted response profiles***</td>
</tr>
<tr>
<td></td>
<td>Validation to reduce field visits*</td>
</tr>
<tr>
<td><strong>PROCESS &amp; ANALYSE DATA</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coding: Adding to indexes/classifications***</td>
</tr>
<tr>
<td></td>
<td>Cleaning and editing*</td>
</tr>
<tr>
<td></td>
<td>Edit &amp; Imputation of single year of age*</td>
</tr>
<tr>
<td></td>
<td>Coverage bias adjustments**</td>
</tr>
<tr>
<td></td>
<td>Placeholders in record imputation*</td>
</tr>
<tr>
<td><strong>OUTPUTS</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>QA collected data, population estimates and characteristics***</td>
</tr>
<tr>
<td></td>
<td>QA detailed matching in areas of greatest uncertainty; CQS triangulation**</td>
</tr>
<tr>
<td></td>
<td>Adjust for collected data in communal establishments**</td>
</tr>
<tr>
<td></td>
<td>Maintenance of Output Areas**</td>
</tr>
<tr>
<td></td>
<td>Journalistic topic analysis***</td>
</tr>
<tr>
<td></td>
<td>Extended output categories eg qualifications***</td>
</tr>
<tr>
<td></td>
<td>Replace previously collected or creation of new variables.**</td>
</tr>
</tbody>
</table>

Key  
- Aggregated data  
- Record level data  
- ***Confirmed use  
- **Likely use (have demonstrated)  
- *Possible use (research still to do)
A. Preparation and collection

14. To maximise response to a census reaching over 25+ million households and ensure efficient use of census resources, some baseline facts and figures are needed for planning and testing. ‘How many…’ ‘how much…’ and ‘where is most effective to…’ type questions need reasonable working estimates so the right amount of census resources can be put in place.

15. We’re researching the potential for alternative data sources to help us answer such questions. Alternative data sources can be valuable in the planning and collection of a census in four main ways.

1. Compilation and validation of the address list

16. A high-quality, authoritative address list is essential to support the census operation, maximise coverage and ensure the appropriate information is sent to both residential households and communal establishments.

17. The core of this register is the National Address Gazetteer (AddressBase). This is maintained by GeoPlace on behalf of Government and incorporates address information supplied by local government, Royal Mail and Ordnance Survey. Address-level intelligence from alternative data sources will be used to assess and where possible improve the quality of the address register. For instance:

   (a) Land registry ‘prices paid’ data for residential properties (to compare property types with those on AddressBase, and other possible uses for example occupation of new builds);

   (b) Number of electricity meters per postcode from Department of Energy & Climate Change (DECC) (possible use, for example to compare number of residential meters in an area against AddressBase’s residential addresses to highlight any anomalies);

   (c) Data from housing websites (such as Zoopla) to identify small areas with concentrations of retirement properties, holiday homes and gated communities.

18. Additionally, alternative data sources will be used to assess the quality of information on communal establishments. This includes open data from Care Quality Commission to assess information on care homes.

2. Sampling – for testing and the Census Coverage Survey

19. Alternative data sources have been used to sample geographic areas for testing of census processes and will underpin the sample design for the 2021 Census Coverage Survey.

20. For some testing we want to test specific outcomes that need specific populations. This includes the testing of the ethnicity question in areas with high concentrations of particular ethnic groups or testing field procedures in particular types of communal establishments.

21. For example, a representative sample of local authority areas were selected to participate in the 2017 Test using:

   (a) Response rates to the 2011 Census;

   (b) 2011 Census data on language, ethnicity, number of students and people aged 65 plus;

   (c) Take-up of broadband connections by postcode and signal strength from Ofcom classified into high, medium and low as a proxy for expected online response rate;
(d) Annual population survey data on sexual identity.

22. More information on choosing test areas and a summary of outcomes is published in (ONS, 2017a).

23. Coverage assessment involves conducting a large-scale independent survey called the Census Coverage Survey (CCS) shortly after the census. This survey will be similar to the survey conducted in 2011 (ONS, 2012b), with its design influenced by alternative data sources. For example, the design of the CCS including the stratification and the sample size will be driven by likely response rates to the census and the CCS.

3. Prioritisation and targeting: Estimating volumes for planning

24. Previous censuses have shown that certain population groups are less likely to complete their census questionnaires for a number of reasons. These are, by their nature, specific and need special attention to encourage their engagement with the census. Historically, these ‘hard to count’ groups include, for example, young men, students, certain ethnic groups, and the very elderly. A primarily online census may mean traditionally compliant groups such as the elderly, who may lack digital skills, become harder to count. Therefore, new groups that cover the digitally excluded will be considered.

25. ONS will use alternative data sources to help identify and prioritise these population groups through the development of a ‘hard to count index’ and other tools, providing timely intelligence about the characteristics of the population. This is so special procedures can be developed to ensure these groups are encouraged and assisted to take part in the census. This will include early engagement with the groups to understand their needs and explore appropriate ways to achieve higher levels of response, use of community advisors, appropriate language support and allocating and prioritising census fieldwork workloads in those areas where the initial return is predicted to be lowest.

26. A critical example here is the prioritisation of fieldwork during the operation based on the characteristics of an area and expected response patterns. This is to help achieve our quality targets of a high overall response rate but also minimising the variation in response between areas. For example,

   (a) Self-response rates from other surveys and previous censuses can help us gauge the numbers and types of people, and their locations, who will need some form of follow-up;

   (b) Smart meter data has the potential to identify communities that are likely to contain higher than average levels of vacant or second residences due to low usage of electricity or water;

   (c) Data from housing websites such as Zoopla can identify small areas likely to contain lots of addresses that are difficult to access (caravan homes used for holidays only, retirement properties and gated communities).

27. Such intelligence would enable field resources to be prioritised, for example to households which are harder to contact or need further encouragement to complete their questionnaire.

28. A further example is what intelligence we can use in alternative sources to identify and target populations who are digitally excluded, to enable them to respond without undue burden. This may include demonstrating where to send paper questionnaires as first contact, as a reminder, or to give out in a field visit. Or sources may indicate populations who would respond online but with additional support. Data on take-up of other online government services and other sources of data on digital inclusion can help us estimate likely locations
and volumes of requests for paper questionnaires, as well as findings from our 2017 Test (ONS, 2017) on the types of people who are likely to need more digital assistance.

29. We are also looking at the outcomes of the 2017 Test to see if there are identifiable population groups who are likely not to self-respond, but do respond to postal reminders. Timely alternative sources would then be used to identify where those groups are likely to be in 2021.

4. Prioritisation and targeting: Re-targeting resources during the live collection operation

(a) Validating field outcomes

30. Once the field operation has started, alternative sources can help validate some of what census officers find out in the field. For instance, if an officer can attempt a visit, but return an assessment that the address:

(a) Doesn’t exist;
(b) Has a different use (holiday accommodation, communal establishment, or commercial property);
(c) Has been demolished or is derelict;
(d) Can’t be accessed;
(e) Is under construction;
(f) Is vacant (unoccupied, or a second address).

31. Timely alternative data can be used to validate these field outcomes: for example, by verifying that an address has become derelict since the address list was created, or on the other hand that council tax is being paid for an address so it is unlikely to be vacant. This will help more effective use of limited field resources by redirecting them during the field operation to those areas most in need of visits, or having more confidence that a response is expected from an address, helping to minimise variability in response rates across areas.

(b) Reacting to response patterns during the collection operation

32. As with previous censuses, we’ll use different strategies to contact non-responding households and encourage them to take part. Alternative data sources will be used before the census to help us understand and model non-response volumes and patterns over the course of the collection period. They’ll also be used to predict the impact of different operational approaches and the most effective timetable for engagement and interventions. We’re working to identify such data and assess ability to predict non-response patterns.

33. During the live collection operation, we will monitor response levels against these predicted profiles and highlight areas or population groups most at risk of not getting sufficient responses, along with an assessment of appropriate interventions (for instance more field visits, reminders, specific community engagement or wider communications initiatives).

B. Data processing and analysis

1. Data capture, cleaning, coding, edit and imputation

34. The increased availability of administrative and other data provides us with the opportunity to improve the quality and timeliness of processing the 2021 Census information. We’ll be exploring how these additional data can be used to help in the various
stages of capturing, coding, cleaning and imputing census data. However, we’ll need to balance whether the improvement in data quality gained from using such data is offset by an increase in processing time.

35. These sources are already being used to maintain and update indexes and classifications, drawing on responses to ongoing surveys to add emerging groups.

36. Alternative data sources could be used is in the edit and imputation stage. As with any self-completion questionnaires, respondents to the census sometimes make mistakes when recording their answers. This results in missing data or invalid responses because they were inconsistent with other answers on the questionnaire. As with previous censuses, we’ll develop an edit and imputation strategy to correct inconsistencies and estimate missing data, whilst preserving the relationships between census characteristics. Alternative sources could be used to providing auxiliary information alongside census data to aid with our imputation model. For example, we’re exploring the feasibility of using administrative sources to improve the imputation of age in the 2021 Census data. Proof of concept work so far indicates this could be promising, but further research is needed to ensure it meets all our quality criteria.

2. Coverage assessment and adjustment

37. Every effort is made to ensure everyone is counted in a census. However, no census is perfect and some people are missed. This under-enumeration doesn’t happen uniformly across all geographical areas or across the population (for example specific sub-groups). We’ll therefore carry out an assessment of the coverage of the census and adjust the data for households and people that have been missed using information from the CCS and the census.

38. Once the number of missing people and households has been identified, we can adjust the data using statistical methods. The approach for 2021 will build on the successful framework and methodology used in the 2011 Census and may involve greater use of administrative and other data. For example, alternative data sources could be used to improve the feasibility of imputing missing household and person records created as a result of coverage assessment, by providing information on the types or sizes of households who at some time have been recorded as resident in non-responding addresses. This builds on the methodology used in 2011, where we used information recorded by field staff on likely accommodation type and number of residents of non-responding addresses.

39. Such an approach depends on the quality and availability of suitable administrative data and the outcome of our work to understand the implications for processing time and the quality gains achieved.

40. In 2011, we used administrative data from large communal establishments to estimate undercoverage of census responses. We expect to follow a similar approach in 2021. We also use alternative sources to detect and correct for biases in the coverage assessment and adjustment process, again building on the 2011 methodology.

3. Quality assuring

41. The 2021 Census results will be subject to a rigorous quality assurance process. As for 2011, this quality assurance process will involve carrying out a series of checks comparing census data to a range of other data sources. Our expectation is that we’ll use a wider range of sources in the 2021 Census.

42. Alternative data sources will help by providing aggregate data to check total numbers of people and households. They can also help with checking demographic comparisons to ensure, for example, the age and sex profiles are as expected. Greater
availability of alternative sources will help us validate the feasibility of census estimates on more census topics, and be used in journalistic-style analysis pieces released with the outputs. In addition, by allowing emerging census estimates to be quality assured dynamically, there are benefits in terms of better targeting of follow-up during the live operation as well as identifying and responding quickly to data anomalies.

43. For instance, for the 2011 census (ONS, 2011), core checks of age and sex profiles used the following comparator sources:

- Birth registration;
- ONS mid-year population estimates;
- NHS patient register;
- Department for Education (DfE) and Welsh Assembly Government (WAG) school census;
- HM Revenue and Customs (HMRC) child benefit data;
- Department for Work and Pensions (DWP) pension claimants data;
- Customer Information Service data from DWP/HMRC on interactions with benefits and taxations systems.

4. Maintaining geographies

44. An area where processing time could potentially be shortened is in the creation of output geographies (Output Areas (OA) and Workplace Zones).

45. Rather than assessing whether the OA boundaries are above or below threshold after adjustment for under-enumeration, we’ll be investigating whether complementary data sources could be used to predict which geographies need to change. We can then adjust them accordingly in advance of the 2021 Census results.

46. The predicted changes would then be compared with final census results and any that don’t meet required thresholds will be adjusted accordingly.

C. Outputs

1. Enhancing census outputs

47. Having collected and processed the census data, this phase involves the analysis, production and dissemination of the census results.

48. For 2021, we intend to integrate census data with other non-survey sources of data to allow us to replace or supplement information collected in the census and create new, enhanced outputs. For example, administrative data could be integrated with census data to provide population characteristics not historically included on the census; one example of this could be integrating information from the Department for Work and Pensions (DWP) and Her Majesty’s Revenue and Customs (HMRC) with census data to produce statistics on household and/or personal income cross-tabulated by census variables such as ethnicity, building on our existing research work in this area (ONS, 2017b).

49. We are currently working to understand the statistical quality of Valuation Office Agency (VOA) data to determine how we could use this information in census operations and outputs. The primary focus of this work is on the potential for these data to replace or enhance the ‘number of rooms’ question on the 2021 Census. We’ll also be evaluating the potential of VOA data to produce new outputs containing information not previously collected in the census, including total floor area of the property and a more detailed
breakdown of property type. In addition to the examples above, we are also planning a programme of research to identify further enhanced outputs based on the administrative and other non-survey data sources available to us.

50. To produce enhanced census outputs, we’ll need to develop and implement methods to bring these data together in a way that would provide statistics of the required quality. Research is already underway on edit and imputation methods; cleaning and matching methods will also be required, as well as a better understanding of how we assess and report on quality of sources, and a strategy of whether these processes are done once for all onward uses, or need to be use-specific.

51. Finally, building on work from the 2011 Census, we will continue to develop methods to use alternative survey and non-survey data to produce output data categories not available directly from the response categories in the collected data (for example, ensuring qualifications categories harmonise to Eurostat output requirements).

2. Quality reporting

52. As in 2011, we’ll give users information about the methods developed to produce the 2021 Census statistics and the quality of the data. We define the quality of a statistical product as its fitness for purpose in terms of the quality dimensions relevance, accuracy, timeliness, accessibility, interpretability and comparability set out in the UNECERecommendations for the 2020 Censuses of Population and Housing (UNECE, 2015).

53. The expected availability of a wider range of data for comparison with the census estimates gives us the opportunity to provide more quality information about the 2021 Census estimates.

V. Opportunities and challenges of using these data in the 2021 Census

54. There are a number of ways in which we believe alternative data sources can help us produce good quality census statistics that meet user needs. This greater use of existing and new data sets has many potential opportunities including:

(a) More cost-effective use of census resources by using intelligence from other data sources to target them effectively;

(b) Improved accuracy of census statistics;

(c) Increased understanding of alternative data sources and how they can be used more widely in the production of statistics;

(d) Improved quality of key administrative data by working in collaboration with the data suppliers to improve data quality;

(e) Production of more timely census statistics than in 2011;

(f) Ability to produce new outputs enhanced to meet user demand for multivariate statistics on subjects not asked on the census, for example income;

(g) Being more responsive to changing user needs by enhancing census data with new sources;

(h) Reduced burden on census respondents to answer questions where alternative data sources are considered to be good enough to replace the census questions, for example on number of rooms (ONS, 2017c);
55. The use of alternative data sources described in this document also provides some challenges. We’ll consider the impact of these as our research progresses. These include:

(a) Their relevance to the statistical concepts being measured in terms of availability of variables and their definitions;

(b) Understanding the processes and procedures associated with the collection, collation, processing and validation of often complex alternative data and the implications these processes may have for the underlying quality of the data;

(c) Access to data sources (particularly microdata) in an appropriate form in the timescales required;

(d) Changes to sources, or the access to them, which could impact on our methods or outputs;

(e) Developing and proving sufficiently robust methods for use of administrative data to provide useful statistics in addition to the 2021 Census;

(f) Balancing increased use of alternative data in process to improve accuracy with the desire to release census statistics more quickly than in 2011;

(g) Additional privacy concerns arising from increased use of person-level data;

(h) Potential circularity of use of the same data in collection and processing but also in QA;

(i) The need to benchmark the Admin Data Census outcome for 2021 against the collected census estimates, given the use already made of the same sources.

56. Considered as a whole, our strategy for the increased inclusion of alternative data in the 2021 Census design has been summarised into the following vision and aims:

**Vision:** to integrate alternative data to provide the best quality 2021 census and leave a positive legacy.

**Aims:** to improve and enhance:

(a) Collection – value for money and optimise response rates;

(b) Processing – improve quality and trust through assurance;

(c) Outputs – improve timeliness and meet more user needs;

(d) Post Census – to continue to add value through what we leave behind.

57. We are formulating design principles around the balancing of cost and gains, building from the Code of Practice principles of Quality, Value and Trust (UK Statistics Authority, 2018), and bearing in mind that just because we can do something, doesn’t mean that we should.

### VI. Conclusions and next steps

60. We are still in the research phase, to determine whether some of these potential uses will meet our inclusion criteria for the 2021 Census design. This phase of development will end in a large-scale rehearsal of the systems and services needed to deliver the census collection, processing, analysis and dissemination operations. We will gather and use as much alternative data sources as possible for the rehearsal. Information and feedback from this phase will enable us to finalise the design for the 2021 Census.
61. As part of our strategy and approach, we have drawn up a list of inclusion criteria to aid the decision-making process, shown in Figure II. This draws on the UK Statistics Authority Code of Practice for Statistics (UK Statistics Authority, 2018) and UNECE Recommendations around quality management (UNECE, 2015), and will be contributing to and drawing on the work of the UNECE Task Force on the measurement of quality of administrative sources for use in censuses (UNECE, 2018).

Figure II
Potential uses of alternative data in the 3 stages of the end-to-end 2021 Census operation

<table>
<thead>
<tr>
<th>INCLUSION CRITERIA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Balance</strong></td>
</tr>
<tr>
<td><strong>Detail</strong></td>
</tr>
<tr>
<td><strong>Quality</strong></td>
</tr>
<tr>
<td><strong>Value</strong></td>
</tr>
<tr>
<td><strong>Trust</strong></td>
</tr>
</tbody>
</table>

62. Having collected the information, this research phase will analyse the data, produce and disseminate the census results, enhanced by administrative data.

References


ONS 2012a – ONS Annual Population Survey
https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/annualpopulationsurveyapsqm

ONS 2012b – 2011 Census Coverage Survey
ONS 2014 – National Statistician’s recommendation on the future of population statistics in England and Wales
https://www.ons.gov.uk/census/censustransformationprogramme/beyond2011censustransformationprogramme/the*censustransformationprogramme/thecensustransformationprogramme/the...commendationfromthenationalstatisticianandchiefexecutiveoftheukstatisticsauthorityandthe
governmentresponse

ONS 2015 – ONS Labour Market Survey
https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemploymenttypes/methodologies/labourforcesurvey/lfsqmi

ONS 2017a – 2017 Test
https://www.ons.gov.uk/census/censustransformationprogramme/progressanddevelopment/test2017test

ONS 2017b – Research Outputs: income from Pay As You Earn (PAYE) and benefits for tax year ending 2016

ONS 2017c – Estimating the number of rooms and bedrooms in the 2021 census in England and Wales: An alternative approach using Valuation Office Agency data
https://www.ons.gov.uk/census/censustransformationprogramme/progressanddevelopment/questiondevelopment/estimatingthenumberofroomsandbedroomsinthe2021censusinenglandandwalesanalternativeapproachusingvaluationofficeagencydata

ONS 2018 – Administrative Data Census Project
https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject

UK Statistics Authority 2014 – Better Statistics, Better Decisions
https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject

UK Statistics Authority 2018 – Code of Practice for Statistics
https://www.statisticsauthority.gov.uk/code-of-practice/

UNECE 2015 – Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing
http://www.unece.org/index.php?id=41398

UNECE 2018 – Task force on measurement of the quality of administrative sources for use in censuses