Office for
National Statistics

Annual assessment
of ONS's progress on the
**Administrative Data Census:
July 2018**

Contents

1   Introduction and vision

In March 2014, the National Statistician recommended that the census in 2021 should be predominantly online, making increased use of administrative data and surveys to both enhance the statistics from the 2021 Census and improve statistics between censuses. The government's response to this recommendation was an ambition that "censuses after 2021 will be conducted using other sources of data".

Decision-makers (central and local government, businesses, charities, community groups and citizens) increasingly need better local data on the size and characteristics of their population to build better services such as transport links, schools, hospitals and housing. They need to understand the changing dynamics of the population nationally and locally, so they can make the best decisions based on that evidence.

In 2021, we'll continue to meet this through the census. However, as the pace of change is increasing, decision-makers need this information much more frequently than every 10 years. In particular, they need to understand the changing nature of our population and migration to and from the UK more regularly and locally.

This is why we have our ambitious programme of work to put administrative data at the heart of the system. By spring 2020, our population and migration statistics system will be based primarily on data that already exist around government, utilising our data-sharing powers through the Digital Economy Act 2017.

This programme does not stop at providing these population data. We've also been researching how to use data such as these to get an up-to-date picture of other important policy areas, for example:

• commuting patterns to better inform transport policy

• educational qualifications and employment to inform skills policy

• income to inform equality and social mobility policy

We've described the work to meet the need for more frequent information as an Administrative Data Census. Until now, we've focused largely on the ability of administrative data and surveys to provide the information traditionally collected through questions on the decennial census. However, integrating data from a range of sources offers much wider potential to improve statistics on a variety of topics. These topics include:

- fuel poverty

- mental health

- debt

- crime

- inequalities

- ageing

- migration

- housing affordability and provision

This will result in better decision-making across government in line with UK Statistics Authority strategy – Better Statistics, Better Decisions (PDF, 1.4MB).

The assessment presented here uses the same criteria and scope as those used for 2016 (PDF, 763KB) and 2017 in terms of our ability to move towards an Administrative Data Census. We'll be reviewing the structure of this assessment and the criteria in the coming year given the wider opportunities from integrating data.

2 Assessment of progress over the past year – main points

This is our third assessment of our progress towards an Administrative Data Census. As with previous assessments, we summarise our current position and where we expect to be in 2023, when we'll make a recommendation on the future of the census.

You can find further information on the evaluation criteria we've used in our previous assessments.

Figure 1 shows this year's assessment alongside the previous assessments. This year we've also added a "next steps" column to note our priorities for taking the work forward.

This year we have:

- developed a methodology for household size and household composition – this has improved our assessment for this topic from Red/Amber to Amber

- worked with data owners to understand their data and specify requirements, primarily for income and migration datasets

- published population estimates to Output Area level

- published an assessment of two methods for producing coverage-adjusted population estimates

- published a wider range of research into producing population characteristics, including new outputs on new mothers' income, internal migration, household composition, ethnicity, small-area income distributions, commuting flows and labour market status

- started research into producing international migration estimates with administrative data at the core

Figure 1.   Current assessment of ONS's ability to move to an Administrative Data Census

Evaluation criteria		2016 assessment	2017 assessment	2018 assessment	Expected assessment by 2023	Next steps towards improving RAG status
Access to data		RED/ AMBER	AMBER	AMBER	AMBER/ GREEN	New migration and ethnicity data sources later this year
Ability to link		AMBER	AMBER	AMBER	GREEN	A framework in place to de-identify and create links across sources in 2019
Ability to meet information needs of users:	Population estimates	AMBER	AMBER	AMBER	GREEN	Admin data at the core of population and migration estimates in 2019
	Households and families estimates	RED/ AMBER	RED/ AMBER	AMBER	AMBER/ GREEN	To be progressed following development in population and migration estimates
	Population and household characteristics	RED/ AMBER	RED/ AMBER	RED/ AMBER	AMBER	Demonstrating potential for multivariate analysis and new topics in 2018-2019
	Housing characteristics	AMBER	AMBER	AMBER	GREEN	Research into overcrowding estimates from admin data in 2018
Acceptability to stakeholders		RED/ AMBER	RED/ AMBER	RED/ AMBER	AMBER/ GREEN	Consultation on new admin data population and migration statistics autumn 2018 to early 2019
Value for money		AMBER	AMBER	AMBER	AMBER/ GREEN	Cost benefit analysis to support recommendation in 2023 about the future of the census

3 Transforming population, migration and social statistics

3.1 A new system of population and migration statistics

At the Office for National Statistics (ONS) Population and Public Policy forum in September 2017, we described how the context for our work is changing. Decision-makers and users are telling us that they need more information on the impact migrants have while they are in the UK. This includes the sectors in which they work, the communities in which they live and the impact on public services such as the NHS and schools. Furthermore, there is a changing policy context, due largely to the UK exiting the European Union (EU). We therefore need to be able to inform decision-makers over the near-term, but also develop data sources to ensure future policy development and implementation is informed by robust evidence.

We therefore have plans, in collaboration with the Government Statistical Service (GSS), to transform the information we produce on migration over the next two years to meet the changing user needs. We'll begin to put administrative data at the core of evidence on migration in 2019. The importance of improving these statistics and the value of the GSS programme of work has been recognised by the Home Affairs Select Committee's report and the government's response.

This is an important element of our plans to transform to a new administrative data-led system for population and migration statistics for England and Wales by spring 2020. By improving migration statistics in this way, we will in turn improve the quality of our population estimates, which underpin a vast range of other information (see Section 3.4).

In autumn 2018, we plan to publish our findings from feasibility research on utilising linked administrative data to provide international migration flows. This will bring together the work on EU and non-EU nationals to explore the feasibility of different approaches to putting administrative data at the core of migration statistics. This work builds and expands upon our research on an Administrative Data Census to move into a better understanding of population and migration using all available sources.

3.2 Proposals for censuses after 2021

In line with the National Statistician's 2014 recommendation and to meet demands from decision-makers for more frequent granular insights into our ever-changing society, we're taking forward work to transform the population

statistics system. We're developing our use of administrative data (information already collected by the public sector) to produce statistics on the population and housing. However, the administrative data is unlikely to provide the whole picture on its own and therefore will be integrated with surveys (and in some cases commercial data).

Firstly, we'll build upon the work outlined previously on population and migration statistics seeking to produce information for a range of geographical areas from national to small-area on:

- the structure of our population (the number and composition of households)

- the size and characteristics of the housing stock

Secondly, we're exploring how integrated administrative, survey and commercial data can provide more frequent and timely information on the characteristics of the population. This includes information on health, the labour market, commuting patterns and educational qualifications. An integrated data approach offers the prospect of producing new more responsive, timely and granular insights of our economy and society. For example, we could look at questions like:

- What's the impact of having children and caring for them on the gender pay gap?

- How great is the link between personal debt and mental health? And do different interventions work for different people?

- What is the energy consumption for similar families in different-quality homes?

- What are the labour market outcomes (employment and earnings) for different ethnic groups with different skills in different parts of the country?

Thirdly, we're rebalancing our business and social surveys to support the administrative data first approach to:

- fill gaps in the administrative data, for example, administrative data are not available on topics such as hours of unpaid care

- help assess (and where necessary adjust for) differences in the definition available from the administrative data and the definition required, in particular, for main measures on population size, migration and number of households

- reduce our reliance on large population and business surveys (where surveys remain the operations will become more efficient by moving from paper and personal interview-based collection to online)

The burden on the public will be reduced by only collecting data once and re-using those data.

The Digital Economy Act 2017 makes a number of amendments to the Statistics and Registration Service Act 2007. These amendments will help the UK Statistics Authority and ONS gather existing data from government departments and some others where they need it to produce statistics.

3.3 Data security and privacy

When we bring data together, we're conscious of the importance of data security and privacy, as we are when directly collecting information from the public and businesses through surveys and the census. ONS has a proven track record of protecting data, built over many decades of running the census for England and Wales and the UK's biggest regular household surveys. In all cases care is taken to ensure that processing of integrated data is in accordance with our legal, data protection and ethical principles and policies.

As we said in last year's assessment, we've adopted the "five safes" framework to safeguard the privacy of individuals. This provides assurance that the data collected from individuals are only used for research and statistical analysis using the following principles:

- safe people (data are handled by people who have been trained and accredited)

- safe projects (data are only used for research projects that deliver clear public benefits)

- safe settings (data are stored in a secure setting)

- safe outputs (all outputs are checked and confirmed as non-disclosive before they are made available)

- safe data (data are de-identified and have names, addresses and any other identifiable variables removed beforehand)

The five safes are designed to maintain public trust and confidence in how we handle personal data in the production of statistics and statistical research.

In Section 4.2, we describe our changed approach to linking data to improve the quality of the links we make while preserving the privacy of the data. This linkage occurs in a way that does not create linked datasets of people or households with personal identifiers. In producing statistics from integrated data, particularly for small populations, we apply the same rigour in data security and privacy as with all official statistics. Strong precautions are taken so that any data made available to the public in any form comply with the confidentiality provisions in the Statistics and Registration Service Act 2007, Digital Economy Act 2017 and the Code of Practice for Statistics. Measures, known as statistical disclosure control, are taken to prevent the release of statistical information that identifies characteristics about an individual person or household.

3.4 Developing a new approach, consultation and decision-making arrangements

Population statistics underpin a wide variety of other statistics (such as unemployment rates) and inform a vast range of decisions. For example, decisions about local services (such as the number of school places or the provision of health services for an ageing population) and decisions about where to site new businesses. Given the importance of producing high-quality population statistics, it would be high risk to move straight to a transformed system without careful development and testing of the new approach. ONS regularly publishes Administrative Data Research Outputs to help with this.

Once we have obtained, studied and understood the administrative data, the data are used to create Research Outputs, which demonstrate the type and quality of outputs that are possible using those data. The first Research Output about the size of the population was published in October 2015. This was followed in 2016 with an update on the size of the population and with new outputs on the number of occupied addresses ("households") and income from Pay As You Earn (PAYE) and benefits. Further updates were published in 2017 and in early 2018 including:

- population estimates (for a range of geographies from Output Areas upwards)

- internal migration estimates

- number and size of households

- household composition

- ethnicity (using combined survey and administrative data)

- gross income distributions down to small areas

- new mothers' income

- commuting flows using anonymised mobile phone data

- labour market status (including more detail on undergraduate students)

Creating and publishing these Research Outputs is enabling decision-makers and users to consider whether the information meets their needs and to provide feedback.

Consultation with, and the views of, end-users and decision-makers will inform our understanding of how well the new system will meet users' needs. The first phase of this will be carried out in autumn 2018 when we consult decision-makers and users on the shape of the new administrative data-based system for population and migration statistics.

The methodologies underpinning the new approach will also be quality-assured by an expert external assurance panel.

In 2023, we will present recommendations to government as to the future of census arrangements, in the light of the progress that has been made.

This work meets the government's requirement, as set out in their response to the National Statistician's recommendation: "Our ambition is that censuses after 2021 will be conducted using other sources of data...In the period up to 2021. The Authority's plans should include ensuring that adequate research into the use of administrative data and surveys is carried out to enable a decision about the future methodology for capturing population and census data."

4   Progress updates on each of the evaluation criteria
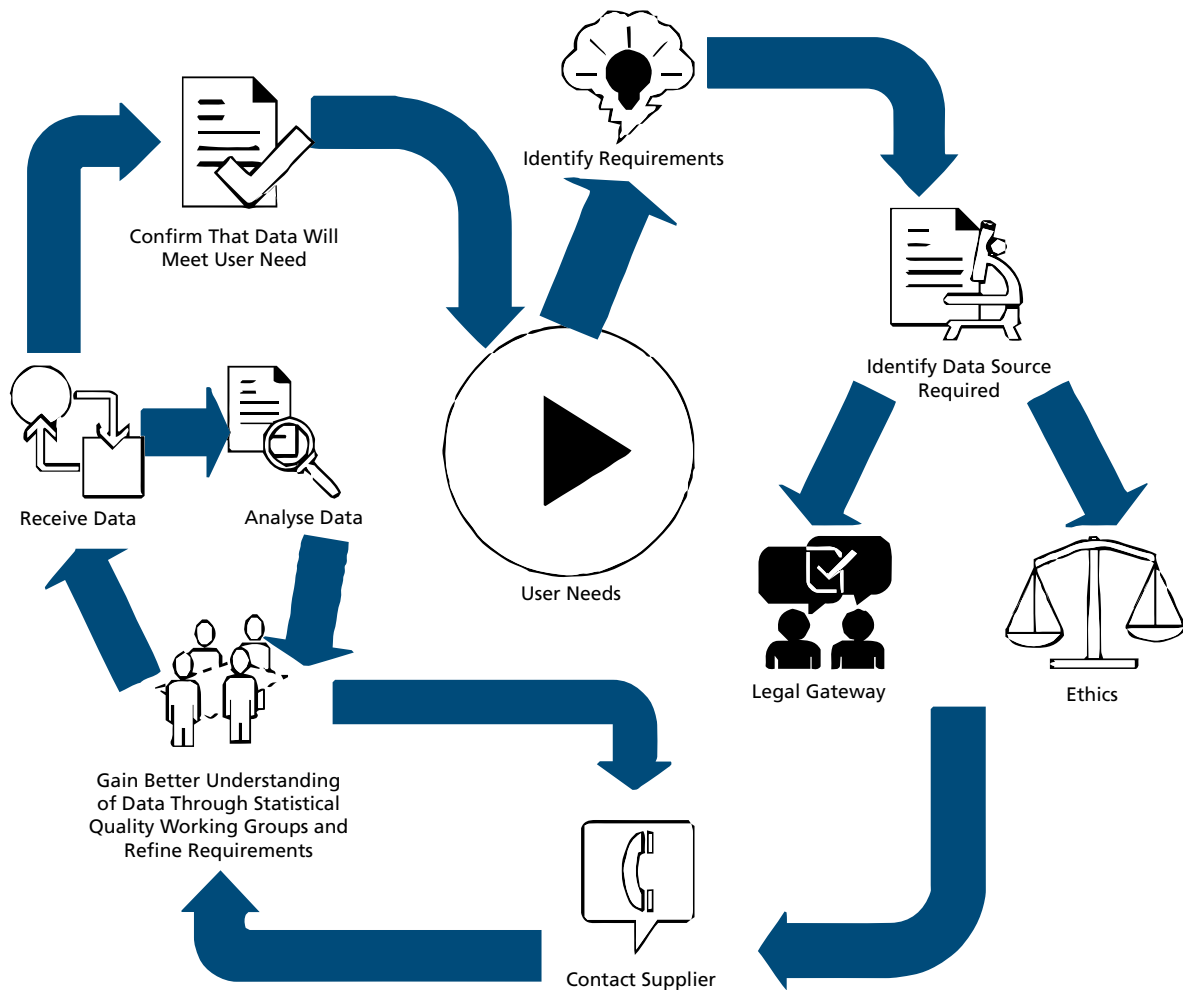
4.1   Access to data

This evaluation criterion is currently assessed to be Amber. Last year, the Digital Economy Act 2017 amended the Statistics and Registration Service Act 2007. The Digital Economy Act 2017 provides a legal gateway for Office for National Statistics (ONS) to access data held by public authorities and commercial undertakings to support the production of official and National Statistics, including the census.

As the codes of practice underpinning the powers of the Digital Economy Act 2017 have been finalised this year, we've begun to collaborate with data providers to establish in detail our requirements for the provision of data. This has included early access to some important new datasets to support our evaluation of progress towards an Administrative Data Census. While we still assess this criterion to be Amber, we expect this to move towards Green/Amber over the coming years as we start to use the new powers of the Digital Economy Act 2017 in earnest.

We've continued to develop our understanding of the data through Statistical Quality Working Groups (SQWGs) where we discuss the data and the detail of the variables with data suppliers such as NHS Digital. They also allow us to understand any changes to the systems and the effect these could have on the data supplied. For example, during the last year, SQWGs helped ONS prepare for the replacement of a previously supplied dataset with an alternative dataset.

The flow diagram in Figure 2 shows, at a high level, the process we go through to receive new data sources.

Figure 2. Process for receiving new data sources

Identify Requirements

Confirm That Data Will
Meet User Need

Identify Data Source
Required

Receive Data

Analyse Data

User Needs

Legal Gateway

Ethics

Gain Better Understanding
of Data Through Statistical
Quality Working Groups and
Refine Requirements

Contact Supplier

This year, we've worked with data suppliers to understand and specify our detailed data requirements for:

• more detailed benefits and income data from the Department for Work and Pensions (DWP)

• Real Time Information Pay As You Earn data and Self Assessment data from HM Revenue and Customs (HMRC)

• health-related datasets from NHS Digital (including hospital episode statistics, mental health datasets and prescriptions)

The next step is the receipt and secure storage of these datasets.

This year, we've accessed new data sources from the Electoral Register, the All Education Dataset for England from the Department for Education (DfE) and an initial feasibility subset of Self Assessment data.

The Electoral Register contains a list of names and addresses, continuously updated through rolling monthly registration, for British, Irish, EU and

Commonwealth citizens aged 16 years or over who have actively registered to vote. This potentially provides access to more current data for this sub-group of the population than other data sources.

The All Education Dataset for England is a longitudinal record-level education dataset that combines the National Pupil Database, and further education and higher education data. It includes a range of characteristics including qualifications. Research continues into the feasibility of using this dataset to provide "Qualifications Held" information.

We've also received administrative data from the Home Office. This extract contains data for non-EU migrants on study, work, family and other visas, which allows us to evaluate immigration and emigration patterns for non-EU citizens.

We're also investigating:

- additional variables in Higher Education Statistics Agency data such as nationality and country of birth

- Tenancy Deposit Protection Scheme and record-level social rented data from the Ministry of Housing, Communities and Local Government and micro-level rental data from the Valuation Office Agency to produce new analyses on tenure

4.2   Ability to link

To create outputs with National Statistics status, we need to use the best available data linkage methods and have a robust understanding of the linkage quality. Until now, we've been linking data using pseudonymised identifiers. This restricts the methods that can be used to link data and makes it more difficult to measure accuracy in a robust and reliable way.

Working closely with our data suppliers, we've moved a number of datasets into our new secure data environment. This environment enables unencrypted identifiers to be used for data linkage, while maintaining security of the data as an ONS priority. This is the approach we referred to in last year's assessment involving the separation of person identifiers (names, dates of birth and addresses) from information about their characteristics. Record linkage can then be undertaken to the highest standards, while still adhering to the principle of not holding identifiable information in one place for longer than required.

This will give us much more flexibility with our methods and allow improved linkage quality. We'll be able to use statistical methods, including probabilistic linking, which are difficult to use confidently with pseudonymised identifiers. Also, it will make the assessment of the quality of the linkage easier and more comprehensive. Better data linkage will improve the accuracy of the outputs

and can reduce bias in the results, since the demographics of people who can be difficult to match are often those of particular interest, for example people who move house frequently.

Over the next year, we'll make this approach our standard. We also intend to evaluate the accuracy of our current methods and further refine them. We currently assess this criterion as Amber, but expect it to move to Green/Amber next year as we fully implement the new approach.

4.3 Ability to meet information needs of decision-makers and users

4.3.1 Population statistics

The population statistics evaluation criteria include both providing mid-year population estimates and components of change. The mid-year population estimates are the stock or count of the population at a point in time. The components of change are the flows between each mid-year estimate (births, deaths, immigration, emigration and internal migration).

Our current assessment is that developments for both the mid-year population estimates and components of change are dependent on us having a population coverage survey live in the field and a methodology in place to use it to adjust our estimates. Given this breadth, we've expanded the criteria to demonstrate the progress across this work. Previously, we've focused on producing stock estimates from our Statistical Population Dataset (SPD) method. This year, we've expanded our research to include components of change, primarily migration. Our current assessment is summarised in Figure 3.

Figure 3. Population and migration estimates assessment

Evaluation criteria	2017 assessment	2018 assessment	Next steps to improve RAG
Population and migration estimates (overall assessment)	AMBER	AMBER	
Annual population base	AMBER/ GREEN	AMBER/ GREEN	More activity data is required to improve the population base
Components of change	RED/ AMBER	RED/ AMBER	New migration data sources are beginning to be available
Coverage survey	RED	RED/ AMBER	Testing if the required response rate can be achieved

Annual population base

For our population base, this year for the first time we've produced population estimates to Output Area level. This used our current Statistical Population Dataset (SPD) methodology approach. We assess these to be Green/Amber. To improve the quality of these estimates, we need a rules-based approach that makes best use of "activity[1]" data and a survey, which enables us to measure and adjust for coverage issues.

More activity data are required to make this possible. In the next year, we expect to have access to activity data from Hospital Episode Statistics, Home Office administrative data and more detailed information from HM Revenue and Customs (HMRC) datasets.

Components of change

This year, we've started to investigate alternative approaches to measuring the size and structure of the population and the components of population change. Decision-makers and other users have told us it is important to understand more about the dynamics of population change. Rates of change are also an important element for producing population projections.

---

1 "Activity" can be defined as an individual interacting with an administrative system, for example, for National Insurance or tax purposes, when claiming a benefit, attending hospital or updating information on government systems in some other way.

We're currently looking at the feasibility of a flows-based approach to producing administrative-based population estimates. This approach is different to the current SPD method. It attempts to use the components of population change (flows) to update continuously the usually resident population rather than producing an independent stock estimate of the population each year.

We've assessed components of change to be Red/Amber. The status reflects that we have already good administrative data on births and deaths registrations, but that considerable challenges remain with identifying administrative data on migration.

We're exploring other data sources that can help us identify different types of international migrants including long-term migrants who are usually resident in the UK. We're interested to learn more about how migrants interact with public services and how combining data sources can tell us about their impact on education and health services and the economy.

This year, we've conducted research into internal migration estimates from SPDs.

The migration statistics transformation update published in May 2018 includes our plans to produce estimates of migration with administrative data at the core. In autumn this year, we plan to publish our findings from the feasibility research on utilising linked administrative data to provide international migration flows. Our research so far suggests that it is much more likely that we'll be able to produce administrative data-based estimates of immigration than emigration.

Coverage surveys

This year, we've completed a small-scale test of a voluntary, mixed-mode Population Coverage Survey to assess response rates for different modes and demographics. The next steps of our research are to test if we can achieve the required response rate alongside testing different sampling strategies.
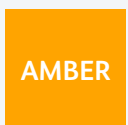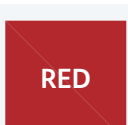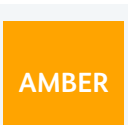
We've also produced a report on coverage-adjusted population estimates to compare two methods. This used an SPD adjusted with activity data in combination with a Population Coverage Survey (PCS) drawn from 2011 Census data to produce coverage-adjusted population estimates for 2011 by local authority. This work evaluated two methods – dual system estimation (DSE) and weighted class estimation (WCE). The results show that there is a higher prevalence of bias in the resulting estimates compared with variance. For most age-sex groups, DSE produced a positive bias (overestimation of the population) and WCE produced a negative bias (underestimation of the population). It is too early to be confident about which of these estimation methods would perform best. More sources of administrative data, including "activity" data, are required to improve the quality of these estimates.

We're currently working with data suppliers including the Home Office, Department for Work and Pensions (DWP), NHS Digital and HMRC to acquire further activity data.

4.3.2 Household and family statistics

We've expanded the criteria into number, size and composition of households to highlight the progress that has been made in each area. Our current assessment is summarised in Figure 4.

Figure 4. Households and families estimates assessment

Evaluation criteria	2017 assessment	2018 assessment	Next steps to improve RAG
Household and families estimates (overall assessment)	RED/ AMBER	AMBER	
Number of households	AMBER	AMBER/ GREEN	More activity data is required to improve these estimates
Size of households	RED	AMBER	More activity data is required to improve these estimates
Composition of households	RED	AMBER	More activity data is required to improve these estimates

This year for the first time we have produced:

- coverage-adjusted household estimates
- household size estimates
- household composition estimates

In July 2017, ONS held a workshop to discuss household statistics, exploring the opportunities and challenges brought about by using administrative data to define households. Over 60 users of household statistics attended from the public and private sector. There were presentations on current household research including the work we've done on occupied addresses, discussions on uses of household statistics and an expert panel session. Discussions emphasised the need for continuous engagement with users in the development of administrative

data-based household statistics and we held a further workshop in March 2018. There were no significant concerns expressed about the occupied address definition and producing statistics from administrative data on that basis.

We also produced a first set of research outputs on household composition which showed more promise than expected, hence the move of this sub-criteria from Red to Amber.

Our publications demonstrate progress in each subcategory of this criteria. As a result, we've improved our assessment of the overall evaluation criteria from Red/Amber to Amber.

4.3.3 Population characteristics

In last year's assessment, we published a diagram outlining the quality and availability of data for producing census topics. This year, we've updated that diagram to demonstrate our progress. See Figure 5. The composite quality measure shows how close the definition of the data is to what's needed by decision-makers and users. It also shows whether there's a known error in the data.

Purple diamonds represent the topics that we've published research on so far. This includes this year's publications on labour market status and ethnicity.

There is a high user need for statistics on labour market status as they reveal areas of deprivation, inequality and labour market exclusion. This subsequently informs policy and economic decision-making at both the national and local level. We've published two Research Outputs this year that demonstrate progress on this topic area – labour market status estimates to local authority level and estimates of students in employment.

In May 2018, we held an event on "Transforming Labour Market Statistics by Integrating Survey and Non-Survey Data" to seek feedback on using administrative data for labour market statistics. This workshop included presenting our work on estimating labour market status from administrative data. At the workshop, there was an understanding of the challenges we face when using administrative data, however, there was also enthusiasm for the extra benefits administrative data could provide including the potential to estimate new areas of the labour market. We received feedback on our methodology and will use that to direct the next steps of our research.
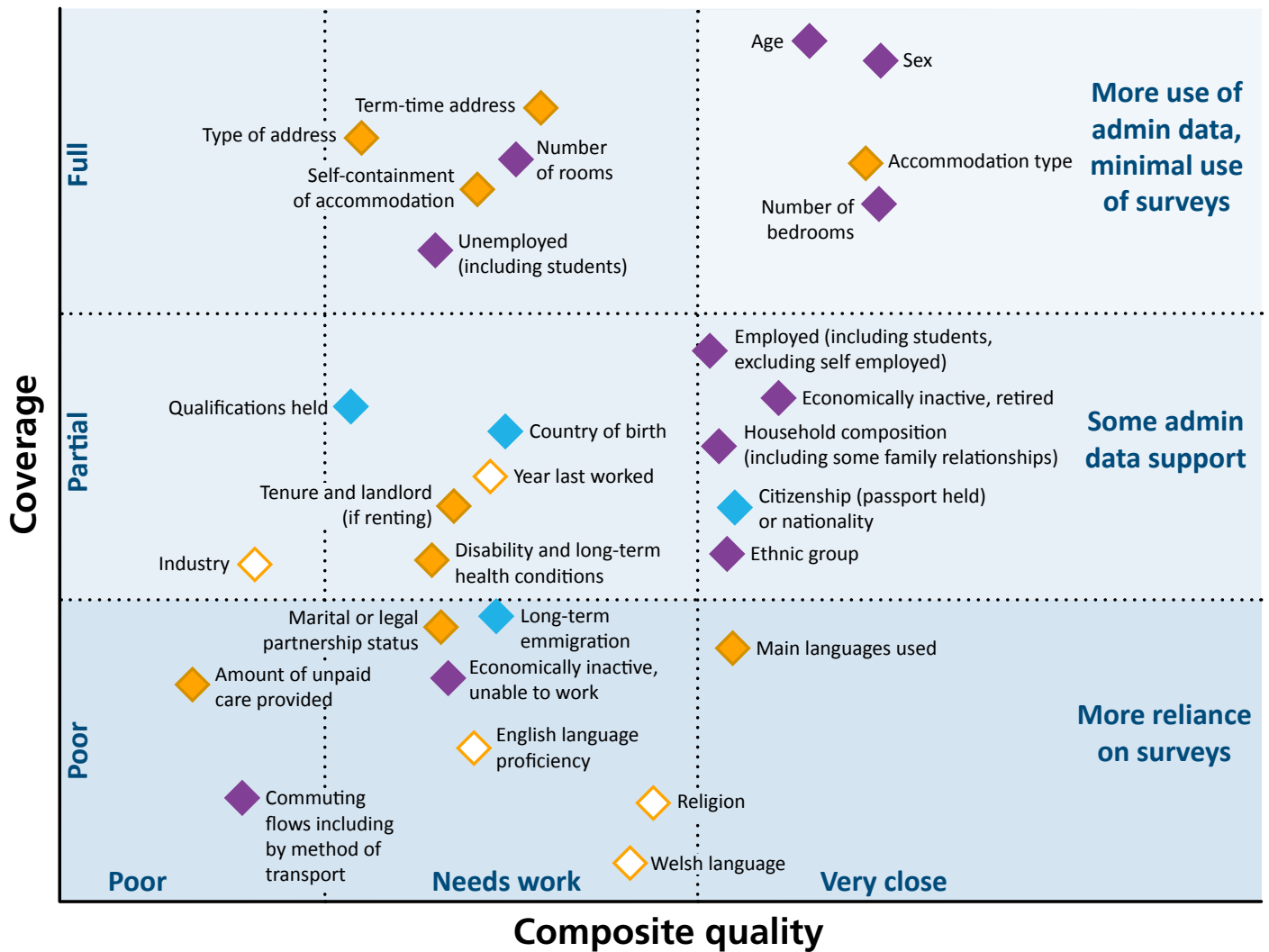
We've also started some feasibility research into the use of machine-learning techniques to predict number of hours worked using survey responses. Number of hours worked is not available in any of the administrative data sources that we currently have access to. We plan to continue developing this research once we can link it to additional data sources.

Ethnicity is a high-priority characteristic as it is an important demographic component of populations. Ethnicity data are used in resource allocation and policy development, and by organisations to monitor and meet their statutory obligations under the Equality Act 2010. This year, we published new ethnicity estimates investigating a new approach combining administrative and survey data.

Using administrative and survey data showed promise for the ability to produce estimates of ethnicity by local authority. However, the survey data are less accurate for smaller ethnicities due to small sample sizes, which means the combined administrative and survey data estimates were also less accurate.

The blue diamonds in Figure 5 represent the census topics we're currently working on and intend to publish research on by the end of 2019. This includes qualifications (now that we have access to data from the Department for Education (DfE) and nationality and country of birth (where we are investigating the use of census data alongside administrative data).

Figure 5. Quality and availability of data

Coverage (vertical axis: Poor, Partial, Full)

Composite quality (horizontal axis: Poor, Needs work, Very close)

Quadrant labels:
- More use of admin data, minimal use of surveys
- Some admin data support
- More reliance on surveys

Plotted characteristics:
- Age
- Sex
- Term-time address
- Type of address
- Number of rooms
- Self-containment of accommodation
- Accommodation type
- Number of bedrooms
- Unemployed (including students)
- Qualifications held
- Country of birth
- Year last worked
- Tenure and landlord (if renting)
- Disability and long-term health conditions
- Industry
- Employed (including students, excluding self employed)
- Economically inactive, retired
- Household composition (including some family relationships)
- Citizenship (passport held) or nationality
- Ethnic group
- Marital or legal partnership status
- Long-term emmigration
- Amount of unpaid care provided
- Economically inactive, unable to work
- Main languages used
- English language proficiency
- Commuting flows including by method of transport
- Religion
- Welsh language

Legend:
- ◆ (purple) Data available to ONS and we have published research
- ◆ (blue) Data available to ONS and we intend to publish in the next 12 months
- ◆ (orange) Some data available to ONS
- ◇ (open) Some data exists, but it is not available to ONS

Characteristics where data is not available, and there is insufficient meta-data to subjectively score:
- Number of hours worked
- Second residences (other addresses)
- Number of cars and vans

Characteristics with no accessible data:
- General health
- National identity
- Supervisor status
- Occupation

NB: Where multiple data sources exist that could be used to derive a characteristic, the assessment is based on the data that is currently available to ONS.

This year, we've also made progress in our income outputs research. Income is a variable that decision-makers and census users repeatedly ask to be included on the census, as it would provide an understanding of deprivation and affluence for small areas. However, it has never been included in UK censuses due to the negative impact on response and poor quality of resulting data. This year, we published progress towards the development of both individual and household income estimates down to Lower layer Super Output Area level.

We also published some research into new mothers' income, improving our understanding of the behaviour of women around the time of childbirth. With additional data, this research could inform the gender pay gap and the provision of services supporting new mothers. This publication also demonstrated the feasibility of combining income information with other data sources to learn about population subgroups.

This year, we also published some analysis using mobile phone data to estimate commuting flows. There are a variety of user needs for these estimates including transport planning and monitoring, the planning of housing and related infrastructure, and labour market and economic planning.

In future years, we'll continue to expand our work on population characteristics and will research additional areas not previously included on the census as described previously in our vision. This research will be driven by priority user needs.

4.3.4 Housing characteristics

This year, we published our assessment of estimating number of rooms and bedrooms using Valuation Office Agency (VOA) admin data. This research evaluated VOA data as an alternative to estimating the number of rooms and bedrooms on the 2021 Census. Lessons learnt about data quality are also applicable to an Administrative Data Census. We held a public consultation, inviting users to respond to this publication. We then published our response evaluating the Valuation Office Agency data.

In the consultation, respondents highlighted some quality concerns with the VOA data. These included the differences between VOA and 2011 Census for occupancy rating (bedrooms) and how frequently VOA records are updated.

We conducted further research to understand these issues in more detail and concluded that, at present, these quality concerns would indeed impact data use. Therefore, we intend to recommend asking number of bedrooms on the 2021 Census.

We'll continue research into understanding the VOA data and how it can be used both to enhance the 2021 Census outputs and in an Administrative Data

Census. This will include research to understand the quality of the number of rooms data and to consider how the property size and type variables could be used.

5 Acceptability to stakeholders

We've continued to consult with our users via a series of local authority engagement events and through the Administrative Data Census contact email. If you would like to get involved, please get in contact with us by email at Admin.Data.Census.Project@ons.gov.uk.

5.1 Engagement activities

This year, we've held two user events about household and labour market statistics. We'll continue this approach focusing on other topics.

Alongside the close-working relationships on the 2021 Census, Office for National Statistics, National Records for Scotland and the Northern Ireland Statistics and Research Agency are also collaborating on the use of administrative data in population statistics and the 2021 Census.

Scotland is also considering future options for enhancing census and population statistics using administrative data and Northern Ireland is considering options for using administrative data to support the 2021 Census.

We continue to engage on the wider international stage with other national statistical institutes on a bilateral basis, through specially-convened working groups and the auspices of the UN. Many other countries are on a similar journey to us in recognising the opportunities in integrating data. Such collaboration is particularly useful in addressing the challenges of using administrative data, which are often common.

We also hosted an Integrated Data for Population Statistics Conference in July 2018 to update stakeholders on our progress and seek feedback.

6 Value for money

At this time, there has been no formal assessment of value for money. In the previous Beyond 2011 Programme, we published a Summary of benefits of census information (PDF, 170KB). We'll be reviewing the benefits in due course to inform the recommendation about the future of the census.

7   Next steps

By the end of 2019, we will:

- review the structure of this assessment and the assessment criteria given the wider opportunities from integrating data

- consult on the shape of the new administrative data-based system on population and migration statistics (autumn 2018)

- consult on needs in readiness to implement a new population and migration statistics system with administrative data at the core (by spring 2020)

- publish our findings from the feasibility research on utilising linked administrative data to provide international migration flows (autumn 2018)

- continue to hold user events on specific topics as appropriate

- pursue the acquisition of:

    - HM Revenue and Customs Self Assessment and further Pay as You Earn (PAYE) data

    - further Home Office administrative data as necessary

    - mental health, Hospital Episode Statistics and prescriptions data to explore potential for new analyses

- apply our improved approach to data linkage in our new secure data processing environment whilst adhering to our legal, privacy, data protection and ethical obligations

- produce Research Outputs on:

    - educational qualifications (by end-2019)

    - income distributions (using PAYE and benefits data) by ethnicity or nationality (late 2018 to early 2019)

    - household composition by income (late 2018 to early 2019)

    - income including Self Assessment data (subject to data access, by end-2019)

    - labour market (subject to access to more detailed Pay As You Earn data, by end-2019)

    - ethnicity for small areas (using integrated census and admin data, by end-2018)

    - overcrowding estimates using Valuation Office Agency (VOA) data and research into use of floor space and property type (by end-2019)