

**Economic and Social Council**Distr.: General
12 July 2023

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses**Twenty-fifth Meeting**

Geneva, 20–22 September 2023

Item 5 of the provisional agenda

The future of population statistics**Statistical transformation of the population and social statistics system in England and Wales****Note by the Office for National Statistics, United Kingdom****Summary*

This document describes the proposed population and social statistics system, built around the Dynamic Population Model (DPM). The current population estimation system in England and Wales relies on the traditional census. This provides granular data every 10 years, but the quality of population estimates declines between census years. The new system will deliver high quality population statistics, every year.

The DPM is a modelling approach which uses various data sources to measure population and components of population change, producing a coherent set of estimates. We have used the DPM to produce admin-based population estimates for mid-year 2011 to 2022 for all local authorities in England and Wales.

For international migration, we replace survey-based methods with administrative data. Rather than survey-based migration intentions, we estimate migration using observed activity. Using the best available data including tax and benefit records, as well as information on travel visas and data at the ports, we publish estimates on immigration, emigration and net migration for non-EU, EU and British nationals..

* Prepared by Louisa Blackwell and Dominic Webber.

Acknowledgements: The Dynamic Population Model has hugely benefitted from close collaboration between demographers and methodologist within ONS. Beyond ONS we have been supported in this research by the University of Southampton. Specifically, we would like to thank John Bryant, Peter Smith, Paul Smith, Jakub Bijak, Jason Hilton, Andrew Hind, Erenkul Dodd and Joanne Ellison for their guidance and support.

Note: The designations employed in this document do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.



The pandemic underlined the need for more timely population estimates. The DPM allows us to publish early, provisional estimates which are confirmed as the data feeds being used mature. In December 2023 we plan to publish experimental 2011–2023 mid-year population estimates, six months ahead of the date scheduled for the publication of the equivalent national statistics.

Flexible by design, our system will be able to reliably adapt to changes in social behaviour, changes in the data available for estimation and changes in users' statistical needs. In this presentation we will also discuss some of the planning and organisational challenges we are encountering as we adapt our statistical processing model to support this increased flexibility.

I. The case for Transformation of the population and social statistics system

1. The census has evolved over time, providing a snapshot every 10 years into who we are and how we live. The census and our census-based mid-year estimates currently provide the best picture of the population at a moment in time. However, the coronavirus (COVID-19) pandemic underlined the need for more timely population estimates and we are committed to maximising the use of administrative data to increase efficiency by making the best use of data already available.

2. We know that our historical census and mid-year estimate approach no longer meets the full range of user needs. We are beginning to transition our methods to a new system that is less reliant on census. Our initial focus is on delivering population and migration estimates at local authority level. We intend to produce outputs that are similar to what users expect currently, but using different methods, which allow the estimates to be timelier and more reliable. Our new dynamic population model (DPM) method is more flexible and will allow us to meet a wider range of user needs. These [experimental statistics](#) give a timely indicator of changes to the population up to 2022.

3. A significant advantage of the DPM is its flexibility. While it uses administrative data sources as stock datasets each year, it can incorporate other data sources when they become available. This could include sources relating to local areas or particular population groups, or sources that represent the total population. The model can also adapt to quality issues in our underlying data sources, drawing strength across data sources and balancing information from population stocks and population flows based on their respective measures of precision. We use a real-time data dashboard with data time series visualisation and outlier detection to monitor recent trends in demographic behaviours and incorporate this intelligence into the model to make it responsive to contemporary trends.

4. Our updated admin-based population estimates (ABPE) have been produced by the DPM despite quality issues in some of the data used for the internal migration component of the estimates. This is a significant benefit compared with our current official mid-year estimates, which [have been delayed](#) in 2023 because of data quality issues.

5. We have defined updated ABPEs as experimental statistics; we continue to refine our methods for producing the ABPEs and these are still subject to further evaluation. Therefore, these statistics have limited use for decision-making. However, the ABPEs demonstrate the potential to produce more timely and coherent estimates of the population compared with our current approaches.

II. A hybrid system for producing timely outputs

6. In Figure 1 we describe the proposed new population and social statistics system. We refer to it as a hybrid system because it integrates different types of data and methods. There is an explicit intention to borrow statistical strength across the different elements of the design, which we now describe. The DPM, at the heart of the hybrid system, both draws other elements in as model inputs, and provides the marginal distributions which, as our best estimates of the population at a given point in time, provide the statistical framework for the other components.

7. *The Dynamic Population Model.* This is described more fully in Section III below. The DPM is a statistical modelling approach to demographic accounting which produces annual mid-year estimates. These are produced as provisional estimates by sex, single year of age and local authority, six months after the mid-year reference point, and as updated estimates 12 months after the reference point. The model uses and automates the cohort component method, updating Census data with the components of population change, but also taking into account information from administrative data about the relevant stock count. Our registration system for births and deaths is robust enough for us to consider birth and death registrations as accurate. A key feature of the DPM is that it combines data with quantified estimates of bias and uncertainty.

8. We consider that the demographic accounts will be as focal to population and social statistics as the National Accounts are to economic statistics.

9. *Longitudinal cohorts.* ONS has been maintaining a 1% Longitudinal Study (LS) which is representative of the population of England and Wales since 1973. The LS links Census data (from 1971 onwards) with administrative data for births, deaths, widow(er)hoods, cancer registrations, immigration and embarkations. This is [a rich research resource](#) which has supported a broad range of research and publications. It has also established, through robust data protocols regarding data access, processing and dissemination, that we can create, maintain and use rich data assets securely and confidently, with appropriate regard to data privacy and protection.

10. We now plan to create the Longitudinal Population Dataset (LDP) which will cover the entire population of England and Wales. The intention is that the LPD will contain key demographic variables and information which will be used to 1) maintain the longitudinal integrity of the cohort and 2) report on its quality and coverage. It will provide the framework for producing satellite longitudinal cohorts, using linked administrative data.

11. We have demonstrated the utility of the approach through the creation of a longitudinal dataset for refugees and asylum-seekers arriving between 2015 and 2020. The [Refugee Integration Outcomes](#) study is already providing unique quantitative evidence of post-arrival integration outcomes in terms of health, geographic mobility and housing circumstances. The LPD will support the rapid creation of longitudinal cohorts, for example cohort studies for all migrants and for armed forces and veterans are planned.

12. The ONS Longitudinal Study will lend statistical strength to the Longitudinal Population Dataset, since record linkage in the former, which is supported clerically, will act as a 'gold standard' comparator for the LPD. Conversely, the LPD will support research studies that the LS cannot; with a 1% sample you can quickly start facing difficulties when doing granular analysis, using detailed classifications or sub-national analysis.

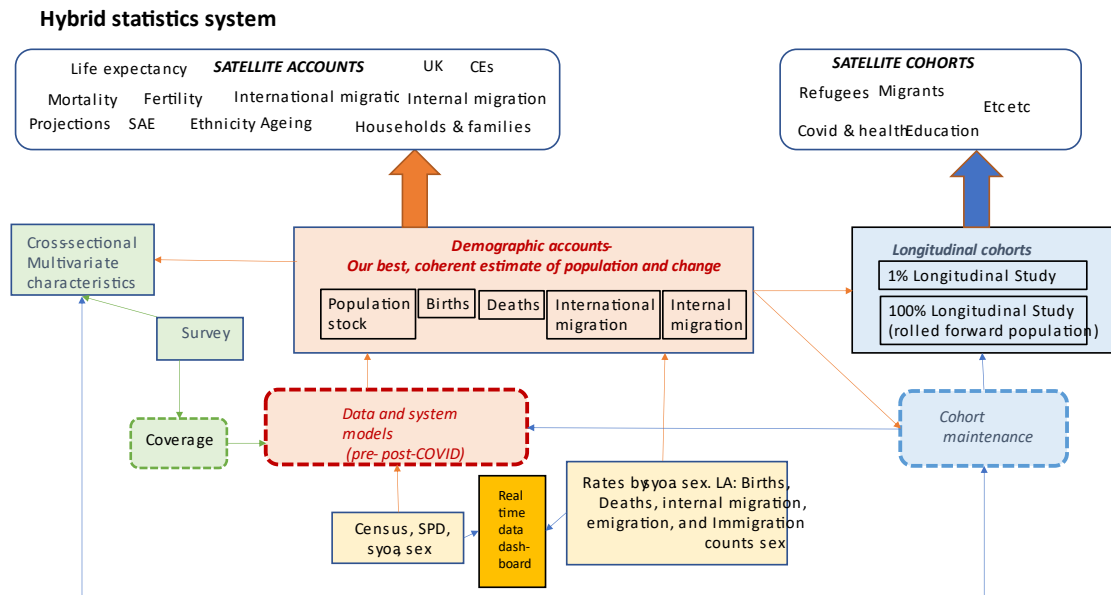
13. The DPM provides ONS' best estimate of the population at a point in time and we will ensure that the LPD is aligned with DPM estimates. We envisage maintaining representativeness of the LPD through weights to account for migration, in the period before administrative sources are mature enough for addition of new migrants' records into the dataset and flagging of departures from the cohort.

14. Conversely, record linkage in the LPD will provide validation checks for the DPM. For example, by comparing birth and death rates generated in the longitudinal LPD data against those derived from data aggregates for the DPM.

15. Survey data has a *key* role to play in the hybrid system. We envisage that our social surveys will support further cross-validation of the LPD. We also need a mechanism for reliably coverage adjusting the statistical population datasets. We envisage using survey data either for coverage adjustment or as an audit survey for our Statistical Population Datasets. We are also committed to providing annual estimates of other population characteristics. Where characteristics are not available in administrative data, we will use social survey data to estimate population characteristics, for the production of annual reports or for attaching characteristics to the longitudinal cohorts.

16. We are currently *focused* on operationalising the DPM to provide annual estimates of population stock and respective flows by sex, single year of age and local authority. We will produce provisional mid-year estimates in December of the reference year and update these in June of the following year. In parallel we will be researching and developing the creation of satellite accounts, covering the broad range of topics currently delivered by existing systems, including household estimates, population projections and topics such as ageing and mortality. There is also high user demand for estimates using definitions other than usual residence. Estimates of population present using different temporal lenses; daytime and night-time populations, seasonal populations for example. Integral to these will be monthly population estimates, capturing seasonality in population flows.

Figure 1
The proposed hybrid population and social statistics system



III. The dynamic population model (DPM)

17. Figure 2 provides an overview of the DPM. Inside the dashed lines are the components of the demographic account that we are estimating. We show that for estimating population stocks, in Census years we have mid-year estimates, which roll forward the usually resident population from Census day (around the end of March) to mid-year (June 30th). In other years we use a combination of Patient Registrations and the Statistical Population Dataset. In each case we include data models, which seek to capture statistical error in these components, usually with reference to the comparison of each source against the censuses. We are currently making the simplifying assumption that our births and deaths data, from our Births and Deaths Registration system, are accurate. For migration, we input smoothed flow rates.

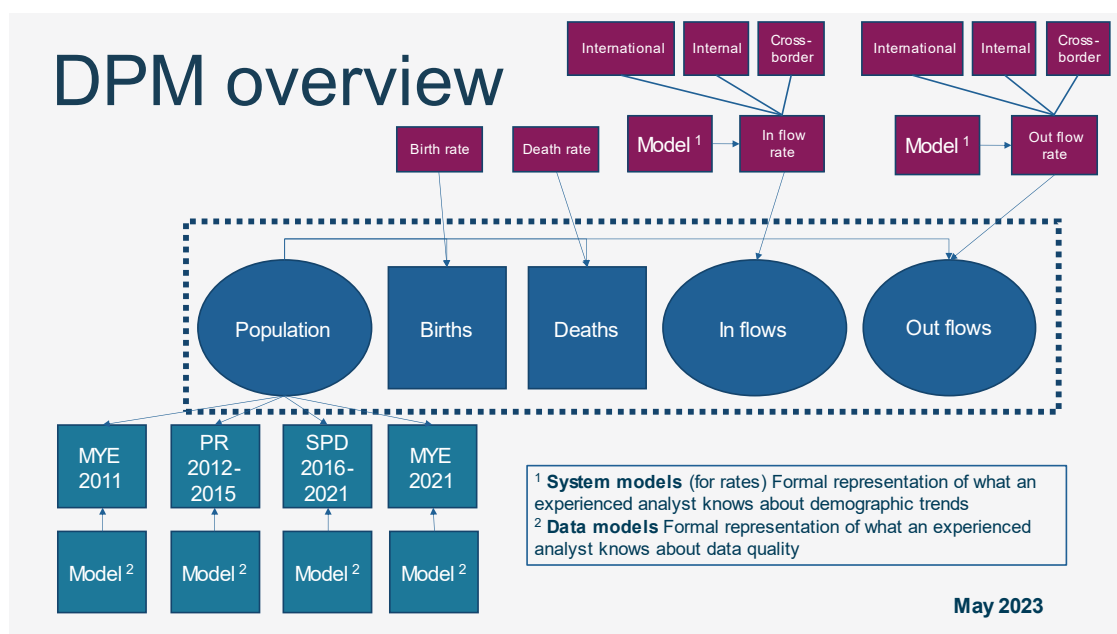
18. We have developed a novel approach to estimating the demographic account, which is fully described in our [technical report](#). We have sped up and simplified the modelling by splitting it into a sequence of stages. Traditional Bayesian modelling involves simultaneously estimating all components and this can be prohibitively slow. In the DPM we create initial

data and system models for births, deaths, migration and population and retain the hyper-parameters from these. We then run the model with hyper-parameters from stage 1 and the real data. This is done separately for each single year of birth cohort in each local authority. We can then re-combine the cohorts and create the demographic account for England and Wales.

19. We have been experimenting with recently-developed techniques for fast, large-scale Bayesian inference. As a result of our reformulation of the model, and the application of the new techniques, computation times have fallen dramatically compared with early versions of the DPM. Our ambition, once the methodology is stabilised, is to open-source the code, along with input data, to support collaboration and further improvement.

Figure 2

Overview of the Dynamic Population Model

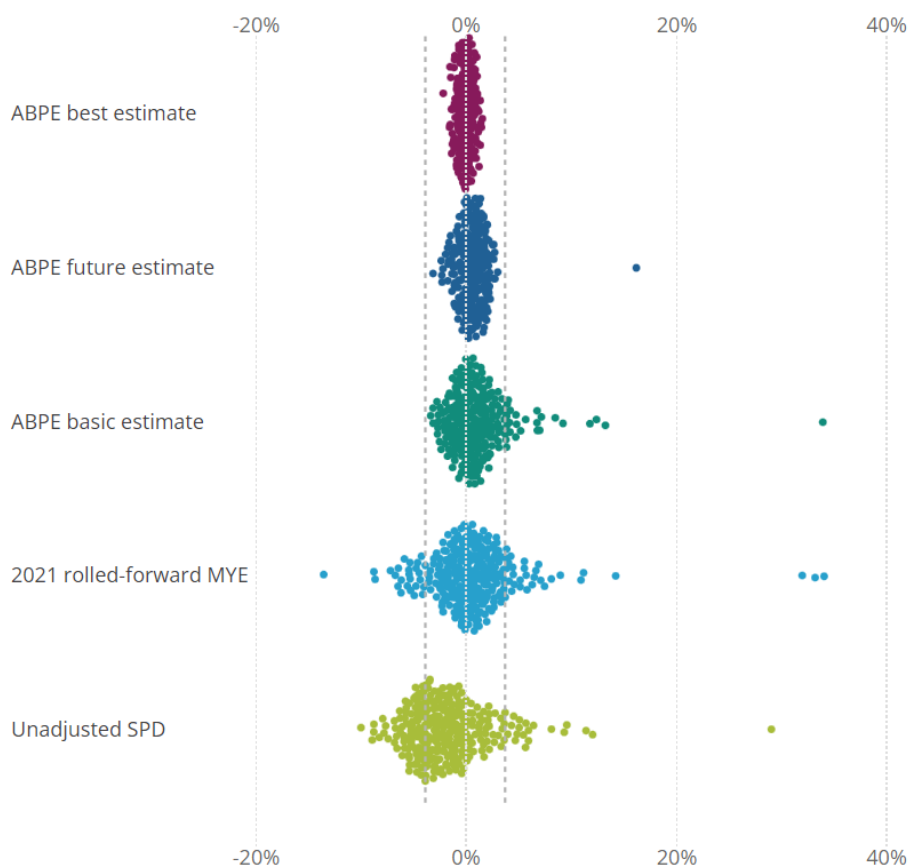


20. We have recently demonstrated the need within the DPM for a robust mechanism to adjust the Statistical Population Dataset (SPD) input to the DPM for under- and over-coverage. This is demonstrated in Figure 3, which shows the total percentage difference by local authority for a range of 2021 admin-based population estimates compared against the Census 2021-based mid-year estimates, for England and Wales. The best estimates from the DPM are those that incorporate the 2021 Census data. Unsurprisingly, these are very close to 2021 mid-year estimates. We show as dashed vertical lines our target quality standard, indicating local authorities falling within 3.8% of the Census 2021-based mid-year estimate.

21. The ABPE 'future estimate' approximates our ambition for a coverage-adjusted SPD to be incorporated into the DPM. Here we use Census data to approximate the coverage adjustment; arguably this is overly aspirational. With a coverage-adjusted SPD we only have one local authority falling outside of our target quality standard. This is Isles of Scilly, with an extremely small population size, is difficult to estimate and even a small absolute difference gives rise to a relatively large percentage difference.

22. The ABPE basic estimate, with no 2021 Census data, and the rolled-forward traditional mid-year population estimates, again with no 2021 Census data and the unadjusted SPDs all have too many local authorities falling outside of our target quality standard. These results underline for us the importance of a robust coverage adjustment for the future system. We are researching how this can be achieved, with the support of survey data.

Figure 3
Comparison of local authority estimates for 2021 using alternative methods



IV. Maximising administrative data use- transformed international migration

23. From 1961 to August 2020, the International Passenger Survey (IPS) was the primary source for measuring international migration to and from the UK. The survey used face-to-face interviews from a sample of passengers to identify migrants as they entered or left the UK, asking how long they expected to stay or leave. IPS data was the largest component of the UK' long-term international migrations (LTIM) estimates until its suspension in March 2020 due to the coronavirus (COVID-19) pandemic.

24. The IPS [was long-known to have limitations in measuring migration](#). For instance, migration estimates were derived from information that respondents provided on their intended length of stay (or departure) in (away from) the UK. Research has shown that for particular subgroups, most notably students, actual length of stay differed from original intentions, leading to inaccurate estimates.

25. This is one of many issues with the IPS that led ONS to launch [the migration statistics transformation programme](#). The programme's main aim was to focus on measuring migration using primarily administrative data – the information collected as individuals interact with public bodies.

26. Since 2019, ONS has progressed research examining the effectiveness of different administrative sources at measuring migration in the UK. Our current approach relies on the use of administrative data from the Home Office (HO) for non-EU visa holders, and the Department for Work and Pensions (DWP) for EU nationals.

A. Measuring migration of non-EU nationals

27. We consider Home Office Exit Checks data – which broadly links information on visa holders with travel data – to be the best source, with the most complete coverage, for measuring non-EU migration. Current methods using HO data use a combination of the “first arrival” and “last departure” of an individual travelling to or from the UK, along with their visa end date, to measure periods of stay/departure for 12 months or more.

B. Measuring migration of EU nationals

28. Until more recently, EU nationals were able to travel to and from the UK freely, without the need of a visa. As a consequence, we’ve needed to develop alternative methods using different sources to those adopted for non-EU. The [current method for estimating the migration of EU nationals](#) uses the Registration and Population Interaction Database (RAPID) dataset. RAPID is a database created by the Department for Work and Pensions. It collates individual’s interactions with the UK tax and benefit system. The current method identifies the total number of weeks of “activity” – i.e. work, or receipt of benefits. Generally speaking, activity that takes place for those who’ve arrived from overseas (or absence of activity) over a 12-month period provides an estimation of immigration (or emigration). Of course, not all people who move to the UK will feature in this dataset. For instance, not all international students will work in the UK while they study, but will be incoming migrants nonetheless. To address these coverage gaps, we use alternative sources to develop adjustments. In the case of students, we link data from the Higher Education Statistics Agency (HESA) to earnings information from the UK tax office to identify the extent to which students have economic activity within a tax year, thereby allowing us to estimate the number of international students not present on RAPID.

29. Currently, we cannot estimate the migration of EU nationals using HO data. This is because of free movement between the EU and UK until January 2021, and continued free movement for EU nationals who have been granted residency through the EU Settlement Scheme (EUSS). With the introduction of new immigration routes for EU nationals through European Economic Area (EEA) visas, these individuals will be present in the HO data. [We are working with the HO](#) to understand if this may allow us to estimate migration of EEA nationals, especially new migrants who are not part of the settlement scheme. This is similar to current methods used for non-EEA nationals.

C. Measuring migration of British nationals

30. Estimating migration of British nationals is more complex. This is because there is no requirement for these individuals to interact with administrative data sources to inform them of an intention to emigrate or subsequently return.

31. Our current methods rely on International Passenger Survey (IPS) data, while alternative sources continue to be explored.

V. The organisational cultural shifts implied by statistical transformation

32. We are transitioning from a Census-dominated model of population estimation to one that is Census-agnostic, making the best use of administrative data. This places much greater reliance on other government departments to supply the administrative sources that we need, reliably and to known quality standards. We are engaging closely with our data suppliers, ensuring that they understand their key role in the production of our population estimation system as we transition to accreditation, over the coming year.

33. Administrative data are not designed for statistical production and are sensitive to policy change. This was seen most dramatically during the Pandemic, for example as engagement with National Health Services changed and created still-to-be-understood changes in data timeliness and quality. Our future system must be resilient to inconsistencies

and interruptions in admin data supplies. For this reason the DPM has been explicitly designed to be flexible, with the ability to incorporate or drop data feeds as required.

34. Our model-based approach to demographic accounting is conceptually more complex than the actuarial-type traditional approach. It is critical that we socialise these new methods in an accessible way. We have a sophisticated dissemination plan that recognises different types of users and their varying appetites for technical complexity. We have been surprised so far by the appetite for technical details that local authority stakeholders have shown and are committed to sharing and building understanding of our methods through teach-ins, which will grow knowledge and confidence not only among users but also within the modelling team. We are producing videos with animated graphics to give citizens, who may not want the technical detail, a true and accessible account of what the model does. Here we are thinking deeply about language, narrative and metaphor to aid communication.

35. The implementation of demographic accounts at scale is new and we are developing the knowledge and capability to operationalise the system. This has involved rapidly upskilling researchers and we are confident that we will develop a strong modelling team capable of extending the methods as the transition progresses; we are containing the current scope to quickly deliver an operational system. In future, we envisage extension to household estimation, small area estimation, monthly estimates and alternative population definitions.

36. Administrative data are subject to change and discontinuity. We are monitoring the data landscape for useful new additions to the DPM; waste water, for example, can support estimation of population present (by sex!). We have become used to stable methods that stand the test of time. In this new approach, flexibility and horizon-scanning are the new 'business as usual'. We will have a continuous research strand running in parallel to the demographic accounting system, seeking enhancements as and when the data allow.

37. There are wider implications of the cultural shift required to support the transitioned system. We need to be more actively engaging with stakeholders and data suppliers. The more local insight and data we feed into the model, the better the estimates will be. This implies a less transactional relationship and a more collaborative approach. We are currently considering, for example, safe and reliable ways to share our model inputs to support collaboration, while previously our data protection procedures were more focused on outputs.

38. We are committed to producing timely statistics and we are thinking creatively about the project plans that will best support this. Rather than thinking of the statistical production process as a long value chain, we are seeking to ensure that by implementing concurrent, parallel development we can deliver outputs faster, with adequate time for quality assurance prior to release.

39. Covid-19 was the catalyst which made automation of the population statistics system an imperative. We plan to open-source the software to support collaboration. We would welcome collaboration with other statistical agencies to help to accelerate and improve these developments.

VI. DPM dissemination strategy-bringing our stakeholders on the transformation journey alongside us.

40. In July 2022 we introduced the [dynamic population model](#) (DPM) as our future proposal for producing timely, coherent population statistics. The previously published admin-based population estimates (ABPEs) were re-branded as Statistical Population Datasets (SPD). This reflects that they are not a finalised estimate, but feed into the DPM where the strengths of the SPD are used alongside other data sources to produce coherent and timely estimates from admin data.

41. In November 2022 we provided [provisional population estimates for 14 case study local authorities](#). In February 2023 we presented the [Admin-based population estimates for local authorities in England and Wales](#) for all 331 local authorities alongside an update to our methods in [Dynamic population model, improvements to data sources and methodology for local authorities, England and Wales: 2011 to 2022](#). In June 2023 we published updated estimates, using further improvements to data sources and methods.

42. In December 2023 we will publish provisional estimates for local authorities, England and Wales, 2011 to 2023. The 2023 estimates from the DPM will be available 6 months earlier than the official mid-year population estimates. By the time we publish updates to these provisional estimates in June 2024, we are hoping to have achieved National Statistics accreditation, making them the official population estimates.

43. We need to bring our users with us through this methodological transition, and appreciate that this methodology is less accessible, conceptually, both for our non-technical audience and key stakeholders. As part of our commitment to transparency and inclusivity, we are producing explainer videos describing how the model works. We are also demonstrating our methods through public engagement events, including scientific conferences and an online teach-in being endorsed by the British Society for Population Studies.
