

[Home \(https://www.gov.uk/\)](https://www.gov.uk/) > [Business and industry \(https://www.gov.uk/business-and-industry\)](https://www.gov.uk/business-and-industry)

> [Charities and social enterprises \(https://www.gov.uk/business-and-industry/charities-and-social-enterprises\)](https://www.gov.uk/business-and-industry/charities-and-social-enterprises)

> [DCMS Sector Economic Estimates: Employment Apr 2021 – Mar 2022 \(https://www.gov.uk/government/statistics/dcms-sector-economic-estimates-employment-apr-2021-mar-2022\)](https://www.gov.uk/government/statistics/dcms-sector-economic-estimates-employment-apr-2021-mar-2022)

Department
for
Digital,
Culture,
Media
&
Sport
(<https://www.gov.uk/government/organisations/department-for-digital-culture-media-sport>)

National statistics

Digital sector workforce analysis - Technical report

Updated 4 November 2022

Contents

1. [Overview of release](#)
2. [Sector definitions](#)
3. [Methodology](#)
4. [Quality assurance](#)
5. [External data sources](#)
6. [Further information](#)



© Crown copyright 2022

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit nationalarchives.gov.uk/doc/open-government-licence/version/3 or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gov.uk.

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

This publication is available at <https://www.gov.uk/government/statistics/dcms-sector-economic-estimates-employment-apr-2021-mar-2022/digital-sector-workforce-analysis-technical-report>

1. Overview of release

The experimental statistics release 'Factors associated with joining or leaving the Digital Workforce' provides estimates of the rate at which people from different demographic groups joined or left the digital workforce between 2012 and 2019, and the impact of various factors on the probability of someone moving into or out of the Digital sector workforce. This is experimental research that builds on and extends previous demographic analysis of employment statistics published as part of the DCMS Economic Estimates series. These estimates are derived from the Annual Population Survey (APS) and contain breakdowns including, but not limited to, ethnicity, region of work, gender and age.

Tables 2 to 4 in the release are based on the Longitudinal APS datasets provided by the Office for National Statistics (ONS). Table 5 is based on the standard APS datasets provided by ONS. More detail on the APS is available in section 3.

The Office for National Statistics (ONS) is the provider of all underlying data used for the analysis presented within this release. As such, the same data sources are used for DCMS estimates as for national estimates, enabling comparisons to be made on a consistent basis.

1.1 Code of Practice for Statistics

DCMS Sector Employment Estimates series is a National Statistic and has been produced to the standards set out in the Code of Practice for Statistics. In June 2019, a suite of DCMS Sector Economic Estimates, including employment estimates, were badged as National Statistics. This affirms that these statistics have met the requirements of the Code of Practice for Statistics. These workforce estimates are a newer set of Experimental Official Statistics that are produced to the standards of the Code of Practice but have not yet been badged.

We encourage users to engage with us so that we can develop our statistics in line with their needs.

1.2 Users

The users of these statistics fall into five broad categories:

- Ministers and other political figures
- Policy and other professionals in DCMS and other Government departments
- Industries and their representative bodies
- Charitable organisations
- Academics

The primary purpose of these experimental statistics is to suggest areas for further research into Digital sector workforce dynamics, helping the Department to understand how current and future policy interventions can be most effective.

2. Sector definitions

In order to measure the size of the economy it is important to be able to define it. DCMS uses a range of definitions based on internal or UK agreed definitions. Definitions are predominantly based on the Standard Industrial Classification 2007 (SIC) codes. This means nationally consistent sources of data can be used and enables international comparisons.

2.1 Digital Sector definition

This section describes the Digital sector definition used in this release, and provides an overview of limitations.

The definition of the Digital sector used by DCMS is based on the OECD definition of the 'information society'. This is a combination of the OECD definition for the "ICT sector" as well as including the definition of the "content and media sector". An overview of the SIC codes included in each of these sectors is available in the [OECD Guide to Measuring the Information Society 2011 \(see Box 7.A1.2 on page 159 and box 7.A1.3 on page 164\) \(https://read.oecd-ilibrary.org/science-and-technology/oecd-guide-to-measuring-the-information-society-2011_9789264113541-en#page166\)](https://read.oecd-ilibrary.org/science-and-technology/oecd-guide-to-measuring-the-information-society-2011_9789264113541-en#page166).

The definition used for the Digital Sector does not allow consideration of the value added of "digital" to the wider economy e.g. in health care or construction. DCMS policy responsibility is for digital across the economy and therefore this is a significant weakness in the current approach.

In addition, there are substantial limitations to the underlying classifications. As the balance and make-up of the economy changes, the SIC, finalised in 2007, is less able to provide the detail for important elements of the UK economy related to DCMS Sectors. The SIC codes used to produce these estimates are a 'best fit', subject to the limitations described above.

3. Methodology

3.1 Data Sources

All analysis in this release is based on Annual Population survey (APS) data provided by the Office for National Statistics (ONS).

Annual Population Survey

The APS is a household survey that combines four quarters of the Labour Force Survey with an additional sample boost. Information collected includes the details of employment (e.g. location, industry, seniority, occupation, income), circumstances (e.g. housing tenure, health) and demography (e.g. nationality, age, ethnicity). Responses are weighted to population totals.

The longitudinal APS combines two waves of survey data from a balanced panel which are weighted to be representative of the UK population and collected 12 months apart. Further details on sample design and weighting can be found in the [Labour Force Survey User guidance](https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyuserguidance) (<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyuserguidance>). Details of respondents' employment are collected during both waves enabling analysis of changes in employment over the period.

3.2 Data processing

The majority of the data processing is done by ONS, with DCMS receiving cleaned and weighted respondent level data. The analysis conducted by DCMS for this report consists of generation of summary statistics, and evaluation of a linear probability model, both of which required the combination of files into multi-year datasets as set out below.

Sample size and dataset combination

The sample size for each year included in the study (Table 1) is insufficient to support intersectional demographic analysis of those joining and leaving the digital sector within single survey years. Only about 3% of the adult (16+) UK population work in the digital sector, and those moving into or out of the sector account for an even smaller proportion of the population. Given, for example, women comprise only 25% of the digital sector, and are relatively evenly distributed across the 5 age bands, an intersectional analysis of gender by age can easily result in unreliable (<30) sample sizes. This particularly affects analysis by ethnicity and disability.

To improve sample sizes, multiple years of data are combined. To avoid combining years with different underlying trends in the data, the data were tested for structural breaks. This was done by calculating the proportion of those employed in the digital industries who left (digital sector leavers) or joined (digital sector joiners) in each study year, and running a Chow test for structural breaks (implemented using the `sctest()` function of the `strucchange` package; Zeileis et al. 2002 (<https://www.jstatsoft.org/article/view/v007i02>)) on the resulting time series. No significant differences in trend for the proportion leaving the digital sector were identified, but a structural break that was not significant at the conventional level ($p = 0.056$) was identified in the proportion who joined in the 2015-16 study year.

Table 1: Sample sizes for each available dataset by population characteristics, UK

Years	Total Sample	Employed in either year	Joined the digital sector	Left the digital sector
2012 - 2013	85,627	56,417	108	94
2013 - 2014	82,962	55,092	96	97
2014 - 2015	81,252	54,699	86	119
2015 - 2016	72,306	48,603	80	95
2016 - 2017	69,318	46,847	84	97
2017 - 2018	65,651	44,714	83	78
2018 - 2019	61,933	42,203	81	83
2019 - 2020	56,018	38,988	68	75

As the Chow test could not be run at the ends of the timeseries, a student's t-test with robust, bootstrapped, standard errors (implemented using `wtd.t.test()` from the survey package; Lumley, 2020 (<https://cran.r-project.org/web/packages/survey/survey.pdf>)) was run to assess the significance of year to year differences in the proportion of digital sector joiners or leavers. There were none (at the 5% confidence level), except between the 2018-19 and the 2019-20 survey years (the proportion of leavers; $p = 0.003$). This is interpreted as the impact of the unprecedented circumstances in 2020 on both data collection and the UK economy, resulting in data that are non-comparable to previous years. The 2019-20 survey year was therefore dropped from the analysis.

For analysis of changes over time and simple summary statistics, where sample size is not as pressing an issue, the potential break at 2016 was incorporated by combining datasets as follows:

- Set 1: 2012 - 13 and 2013 - 14
- Set 2: 2014 - 15 and 2015 - 16
- Set 3: 2016 - 17, 2017 - 18, and 2018 - 19

For analysis of the interaction between factors, where sample size is of more importance, all datasets with the exception of 2019-20 are combined.

Disclosure control

As part of the production process we also apply disclosure control measures to prevent the identification of any respondents. We suppress values where the number of respondents for a cell is below a set threshold. Where appropriate, we also apply secondary suppression to prevent disclosure via differencing.

3.3 Analysis

This experimental analysis was conducted to suggest avenues for further policy and analytical work by estimating whether people in some groups are less likely to join, or more likely to leave the digital sector, and how factors affecting those decisions may interact. This is assessed in two ways: calculation of summary statistics, and evaluation of a linear probability model to account for the effect of interactions between non-independent factors, e.g. age and likelihood of having a degree.

In both cases the characteristics of those joining or leaving digital sector jobs are compared with a baseline population. This is because we may expect that there are omitted variables affecting a person's willingness to change jobs and their success in doing so. To minimise the effect of these omitted variables on the analysis the comparison groups were selected to be as similar to the treatment population i.e., those workers that joined or left the digital sector, as possible.

The comparison group for those joining the digital sector was all respondents who had changed employer and industry (here taken as the standard industrial classification division), but who had not moved into the digital sector. The comparison group for those leaving the digital sector was all those who had changed employer but remained within the digital sector. In both cases, the comparison group was chosen to minimise the effect of unobserved characteristics likely to affect someone's willingness or ability to change jobs (e.g. lack of dependents). In the case of digital sector leavers, comparison to those who changed employer but remained in the sector should also minimise the effect of omitted but observable characteristics, like subject studied, on the analysis.

Summary statistics

In this publication, a person's employment is taken as their main job. In other words, if a respondent's primary employment is not in the digital sector, but they have a second job within the digital sector, they are not considered to be employed in the digital sector for the purpose of this analysis. This is a different approach to that taken for the main DCMS Sector Economic Estimates: Employment publication, which is a measure of the number of filled jobs.

The summary statistics reported are the percentage of the estimated population of interest (e.g., those leaving the digital sector) that have the specified characteristic. The significance of any differences between the demographic profile of those moving into or out of the digital sector, and that of their respective comparison group is assessed using a student t-test with robust standard errors (Lumley, 2020).

The significance is reported as the level at which the difference is significant. In other words a reported significance level of 5% means that the probability of the test wrongly indicating that the difference between the two estimated proportions represents a real difference is lower than 5%.

Linear probability model

The purpose of the model is to provide an initial assessment of factors affecting the probability of a person entering or leaving the digital sector workforce, and their interaction. A linear probability model with some non-linear regression terms was chosen as the output is readily interpretable as the marginal effect, in percentage points, of each regressor on the outcome.

Regressors were chosen based on the known and persistent participation differences between different demographic groups (DCMS, 2021 (<https://www.gov.uk/government/statistics/dcms-sector-national-economic-estimates-2011-to-2020>)), and include gender (S_i), age (A_i), ethnicity (E_{ij}), disability status (D_i), level of education (Q_i), and location (R_i). As it is possible that age and likelihood of having a degree interact due to the expansion of higher education in the UK (ONS, 2017 (<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/graduatesintheuklabourmarket/2017#steady-increase-in-the-number-of-graduates-in-the-uk-over-the-past-decade>))), the model also includes an age and qualification interaction.

Analysis of UK Time Use Survey data (Henz, 2020 (<http://eprints.lse.ac.uk/105796/>)) showed that in families with young (aged < 3 years) children, women provide the majority of unpaid childcare, which can impact their choices regarding paid work. To disentangle the impact of maternity care (or lack thereof) from other aspects of the digital sector work culture that may have a discouraging effect on female participation, we include a dummy variable for the presence of children under 3 years old, i.e. those for whom there is no state supported childcare, in the household (C_i) as well as an interaction term with gender.

Information on the quality and quantity of work, which may motivate job seeking, is limited in the available datasets. These are partially addressed in the regression equation by inclusion of a change in total hours worked (ΔH_i) term, provision of training in the two roles (T_{ij}), and a looking for work dummy variable (L_i). To minimise the number of dummy variables required, and address the non-linear response, age is included as a continuous, de-meaned, squared term in the regression rather than as age bands. An interaction term with gender is included to assess gendered differences in experience and, by inference, seniority. The regression equation estimating the probability of moving into or out of the digital sector (P_i) is therefore of the form:

$$P_i = \beta_{0i} + \beta_{1i}A_i + \beta_{2i}(A_i - \bar{A})^2 + \beta_{3ij}E_{ij} + \beta_{4i}S_i + \beta_{5i}S_i(A_i - \bar{A})^2 + \beta_{6i}Q_i + \beta_{7i}Q_iA_i + \beta_{8i}D_i + \beta_{9i}C_i + \beta_{10i}C_iS_i + \beta_{11i}R_i + \beta_{12i}\Delta H_i + \beta_{13i}L_i + \beta_{14i}T_{ij}$$

To avoid the dummy variable trap, the dummy variable for the most common value of any particular characteristic was dropped from the equation. Testing for multicollinearity found no relations of concern between the regressors.

It is likely that there are several omitted variables, like ambition or non-child related caring responsibilities, that contribute to a person's likelihood of seeking, and obtaining, a new job. This is partially accounted for in two ways: inclusion of the

'looking for work' term in the regression equation, and selection of appropriate comparison groups. Use of an instrumental variable to address the omitted information was rejected due to lack of viable candidates in the available data sets.

As the outcome is binary (whether the respondent joined or left the digital sector), it is expected that the error term will be heteroscedastic. As this would affect the consistency of standard errors and estimation of the significance we use generalised least squares regression, implemented via the `glsl()` function in the nlme package (Pinheiro et al. 2022 (<https://cran.r-project.org/web/packages/nlme/index.html>)). All analyses were performed using R Statistical Software (v4.0.3; R Core Team 2020).

4. Quality assurance

This document summarises the quality assurance processes applied during the production of these statistics by our data providers, the Office for National Statistics (ONS), as well as those applied by DCMS.

4.1 Quality assurance processes at ONS

Quality assurance at ONS takes place at a number of stages. The various stages and the processes in place to ensure quality for the data sources are outlined below. It is worth noting that information presented here on data sources are taken from the [Annual Population Survey \(QMI\)](https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/annualpopulationsurveyapsqmi) (<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/annualpopulationsurveyapsqmi>). This work should be credited to colleagues at the ONS.

ONS Annual Population Survey

The purpose of the APS is to provide information on important social and socio-economic variables at local levels. The APS is not a stand-alone survey, but uses data from the Labour Force Survey (LFS) and a local sample boost.

Sample design

The APS survey year is divided into quarters of 13 weeks. From January 2006, it has been conducted on the basis of calendar quarters: January to March (Quarter 1), April to June (Quarter 2), July to September (Quarter 3) and October to December (Quarter 4). The APS design is not stratified.

The APS data set is created by taking waves 1 and 5 from four successive quarters, with rolling-year data from the English, Welsh and Scottish Local Labour Force Survey, to obtain an annually representative sample of around 80,000 households. Over the period of the 4 quarters, waves 1 and 5 will never contain the same households to avoid the inclusion of responses from any household more than once in the dataset.

Sampling frame

The sampling frame for the survey in Great Britain is the Royal Mail Postcode Address File (PAF) and the National Health Service (NHS) communal accommodation list. Due to the very low population density in the far north of Scotland (north of the Caledonian Canal), telephone directories are used as sampling frames. A systematic sample is drawn each quarter from these three sampling bases, and as the PAF is broken down geographically, the systematic sampling ensures that the sample is representative at regional level. In Northern Ireland, the Rating and Valuation Lists (which serve for the administration of land taxes) are used.

Data collection

Interviews in all waves are carried out either on a face-to-face basis with the help of laptops, known as Computer Assisted Personal Interviews (CAPI) or on the telephone, known as Computer Assisted Telephone Interviews (CATI). Information is collected using a software package called Blaise.

Validation and quality assurance

- Accuracy is the degree of closeness between an estimate and the true value. As both surveys are sample surveys, they provide estimates of population characteristics rather than exact measures. At ONS, confidence intervals are used to present the sampling variability of the survey. For example, with a 95% confidence interval, it is expected that in 95% of survey samples, the resulting confidence interval will contain the true value that would be obtained by surveying the whole population.
- Comparability is the degree to which data can be compared over time and domain, coherence is the degree to which data are derived from different sources or methods but refer to the same topic and are similar. Some sources provide data that overlap with APS/LFS data on employment, unemployment and earnings. More information on these sources are available [here](https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/annualpopulationsurveyapsqmi) (<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/annualpopulationsurveyapsqmi>).
- [Statistical disclosure control](https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/disclosurecontrol) (<https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/disclosurecontrol>) methodology is also applied to the datasets before release. This ensures that information attributable to an individual is not disclosed.
- On each quarterly LFS dataset, the variable frequencies are compared with the previous period. This identifies any significant discontinuities at an early stage. All discontinuities judged significant are investigated to determine the reason for the discontinuity. Is it the product of questionnaire revision or processing error, derived variable revision or error or real-world change? This process also ensures that the metadata associated with each variable are correct.

- Specific main derived variables are checked in detail by extracting the underlying variables and recalculating in another application, then comparing the results with the values in the dataset. This ensures that the program used to calculate the derived variables is working correctly.

4.2 Quality assurance processes at DCMS

The majority of quality assurance of the data underpinning this release takes place at ONS, through the processes described above. However, further quality assurance checks are carried out within DCMS at various stages.

Production of the report is typically carried out by one member of staff, whilst quality assurance is completed by at least one other, to ensure an independent evaluation of the work.

Data requirements

For APS data, DCMS discusses its data requirements with ONS and these are formalised as a Data Access Agreement (DAA). The DAA covers which data are required, the purpose of the data, and the conditions under which ONS provide the data. Discussions of requirements and purpose with ONS improve the understanding of the data at DCMS, helping us to ensure we receive the correct data and use it appropriately.

Production and data analysis

At the production stage, data are aggregated up to produce information about DCMS sectors and sub-sectors before inputting the data into the formal data tables ready for analysis. Disclosure control is also applied as part of this process.

The statistical lead ensures a number of quality assurance checks are undertaken during this process. Where relevant these checks typically include:

- whether disaggregations sum to the overall total.
- “Sense checks” of the data. E.g.:
 - Are the estimates similar from one year to the next? How do the figures compare with ONS published totals?
 - Looking at any large differences between the data and possible causes to these.
- Checking that the correct SIC codes have been aggregated together to form digital sector. Are all SIC codes we require included? Are there any non-digital SIC codes that have been included by accident?
- Checking it is not possible to derive disclosive data from the figures that will be published.
- Making sure the correct data has been pasted to the final tables for publication, are accessible, formatted correctly, and have appropriate documentation.

Having checked the quality of the data, analysis is then conducted to outline the key trends and patterns. This is then checked to ensure all statements, figures and charts are correct.

Model performance

One measure of model performance is to assess the predicted values against the values that should occur if the model predicted outcomes perfectly.

The linear probability model predicts the likelihood of joining or leaving the digital sector relative to a comparison group. If the model predicted the outcomes perfectly, the predicted values would be 100% for digital sector joiners or leavers, and 0% for the respective comparison groups. The average value for digital sector joiners was 4%, and for the comparison group, 2%. The average value for digital sector leavers was 58%, and for the comparison group 52%. In other words, the factors examined are important but do not determine the majority of movements into and out of the digital sector.

Dissemination

Finalised figures are published as OpenDocument spreadsheets on GOV.UK, with summary text on the webpage. These are produced by the lead statistician who, beforehand, checks with the ONS on details of how to interpret the statistics. Before publishing, a quality assurer checks the figures match between the tables and the GOV.UK page summary. The quality assurer also makes sure any statements made about the figures (e.g. regarding trends) are correct according to the analysis and checks spelling or grammar errors.

5. External data sources

It is recognised that there are always different ways to define sectors, but their relevance depends on what they are needed for. Government generally favours classification systems which are

- rigorously measured,
- internationally comparable,
- nationally consistent, and
- ideally applicable to specific policy interventions.

These are the main reasons for DCMS constructing sector classifications from Standard Industrial Classification (SIC) codes. However, DCMS accepts that there are limitations with this approach and alternative definitions can be useful where a policy-relevant grouping of businesses crosses existing Standard Industrial Classification (SIC) codes. DCMS is aware of other estimates relevant to DCMS sectors. These estimates use various methods and data sources, and can be useful for serving

several purposes, e.g. monitoring progress under specific policy themes such as community health or the environment, or measuring activities subsumed across a range of SIC's.

The ONS use the quarterly Labour Force Survey for their estimates of UK-wide employment rates. Our APS employment estimates of the number of filled jobs in the DCMS sectors takes a similar approach. However, as the APS uses two waves of the LFS, the datasets are not directly comparable and the ONS published figures will differ slightly from ours.

For employment statistics more broadly, the main alternative is the Business Register and Employment Survey (BRES). This has the advantage of asking businesses directly about their employees and hence is likely to capture the sector of employees more accurately than a household survey. However, it does not contain the range of demographic breakdowns that the APS does, which enables us to build a fuller picture of employment in our sectors, using a still-robust data source, and does not include the self-employed, which constitute a substantial part of the workforce for many DCMS sectors.

It is recognised that there will be other sources of evidence from industry bodies, for example, which have not been included above. We encourage statistics producers within DCMS sectors who have not been referenced to contact the economic estimates team at evidence@dcms.gov.uk.

6. Further information

For enquiries on this release, please email evidence@dcms.gov.uk.

For general enquiries contact:

Department for Digital, Culture, Media and Sport
100 Parliament Street London
SW1A 2BQ

Telephone: 020 7211 6000

DCMS statisticians can be followed on Twitter via [@DCMSInsight](https://twitter.com/DCMSInsight) (<https://twitter.com/DCMSInsight>).

This release is an Experimental Official Statistics publication and has been produced to the standards set out in the [Code of Practice for Statistics](https://code.statisticsauthority.gov.uk/) (<https://code.statisticsauthority.gov.uk/>).

[↑ Back to top](#)

OGL

All content is available under the [Open Government Licence v3.0](#), except where otherwise stated

© Crown copyright