Department for Digital, Culture, Media & Sport

# Digital Sector Economic Estimates: Earnings Quality Assurance Report

## Contents

## Introduction

This document summarises the quality assurance processes applied during production of the **Digital Sector Economic Estimates 2018: Earnings** release (published 4th September 2019). This release presents analysis on median annual earnings for Digital Sector employees as well as those in Digital Occupations and is based on the Annual Survey of Hours and Earnings (ASHE) dataset provided by the Office for National Statistics (ONS)[1]. This data is the most detailed and comprehensive source of earnings information in the UK.

This document covers quality assurance carried out by both DCMS and our data providers (ONS).

Please note that these statistics are classed as experimental statistics as it is the first time DCMS have introduced analysis on earnings in the Digital Sector as well as Digital Occupations. DCMS plan to widen this analysis further in the future to include all DCMS sectors, if there is sufficient interest in these statistics.

## Quality assurance processes at ONS

The data underpinning this release are taken from the Office for National Statistics (ONS) Annual Survey of Hours and Earnings (ASHE)[2]. ASHE provides information about the levels, distribution and make-up of earnings and paid hours worked for employees in all industries and occupations.

Quality assurance at ONS takes place at a number of stages. The various stages and the processes in place to ensure quality at each stage are outlined below. This information is taken from the ASHE quality information report[3] and should be credited to the ONS.

***Sampling and data collection***

ASHE is based on a 1% sample of employee jobs taken from HM Revenue and Customs (HMRC***) Pay As You Earn (PAYE) records.*** The sample is matched against the ***ONS's*** Inter-Departmental Business Register (IDBR) in order to obtain contact and address details for the employers. Information on the hours paid and earnings of employees is obtained from employers and treated confidentially.

The sample is drawn in such a way that many of the same individuals are included from year to year, thereby allowing longitudinal analysis of the data. Please note that ASHE does not cover the self-employed nor does it cover employees not paid during the reference period.

A specific date in April is chosen so that all respondents refer to the same point in time. This reference date is not the same every year. Given the survey reference date in April, the survey does not fully cover certain types of seasonal work, for example, employees taken on for only summer or winter work.

---

[1] https://www.ons.gov.uk/

[2] https://www.ons.gov.uk/surveys/informationforbusinesses/businesssurveys/annualsurveyofhoursandearningsashe

[3] https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/methodologies/annualsurveyofhoursandearningslowpayandannualsurveyofhoursandearningspensionresultsqmi

A copy of the questionnaire of the survey is available online at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/methodologies/annualsurveyofhoursandearningsashemethodologyandguidance. This also comes with detailed instructions on how to complete (and return) them.

***Response rate and imputation***

The ASHE dataset contains information on approximately 180,000 jobs in all industries, occupations and regions, making it the most comprehensive source of earnings information in the UK and enabling a vast range of analyses.

Sometimes respondents return forms with only some of the variables completed. This means that different variables have different levels of response within the returned dataset. In turn this would mean that the proportions of each variable returned for each job type are different and therefore the weights required for each variable would be different.

Instead of using different weights for different variables we have imputed for the missing values so that the same weights can be used for all the variables.

The imputation method used for ASHE is donor imputation. To impute for a missing value in a record, this method looks for another record with similar characteristics and uses the value from the donor record for the missing variable. This ensures that the distribution of each variable within the imputed data set is similar to the distribution in the un-imputed dataset.

***Weighting***

Weighting is used to compensate for types of job that are under represented in the ASHE dataset due to poor response. Because some types of respondents are more or less likely to respond than others, the data is not always representative of the general population. Therefore different weights are applied to different types of respondents, so that when the weights are added together, each type will have a proportion consistent with their proportions in the general population.

For example, if respondents of type A are generally poor responders, each record of type A will need a relatively large weight to so that they can collectively represent all the type A people in the population.

The general population proportions used by ASHE to calculate its weights come from the Labour Force Survey (LFS) and the types are determined by classifying people by age group, sex, occupation and a regional split. E.g. estimates of the level of earnings for people working in London are increased more than estimates for other regions. This widens the estimate of the difference in pay between London and other regions of the UK.

***Disclosure***

Statistical disclosure control is applied to all outputs produced from ASHE. This ensures that information attributable to an individual or individual organisation is not identifiable in any published outputs. The Code of Practice for Statistics and, specifically, the pillar on trustworthiness, sets out principles for how we protect the identities of respondents from being disclosed.

Firstly, to protect individual earnings data, a frequency count is taken and all cells that are based on a count of fewer than three individuals are suppressed. Secondly, to protect employers' pay information, a dominance rule is applied within each cell, which uses the contribution from the largest employer and the overall standard error of the estimate to deduce whether information about the employer can be derived with a reasonable degree of certainty.

Given the nature and complexity of ASHE outputs it is not possible to use a practical method to check for issues of secondary suppression. Instead, ASHE applies a policy where no sample counts are released, only weighted sample counts rounded to the nearest 1,000. This gives users enough information about the sample size for a cell for them to make quality inferences, without giving sufficient information to derive data by difference with any degree of certainty. Although in some circumstances a figure can be derived by difference, it would be impossible to tell how many individuals contributed to the figure.

***Sampling errors***

This occurs because estimates are based on a sample rather than a census. ASHE estimates this error through coefficients of variation (cv) which are published alongside all ASHE outputs. The cv is the ratio of the standard error (se) of an estimate to the estimate itself, expressed as a percentage. Generally, if all other factors are constant, the smaller the cv the higher the quality of the estimate.

It should be noted that at low levels of disaggregation, high coefficients of variation imply estimates of low quality. For example, for an estimate of £400 with a cv of 10%, the true value is likely to lie between £321.60 and £478.40. This range is given by the estimate plus or minus 1.96 multiplied by the se. Where these ranges for different estimates overlap, interpretation of differences between the relevant domains becomes more difficult.

***Non-sampling errors***

ASHE statistics are also subject to non-sampling errors. For example, there are known differences between the coverage of the ASHE sample and the target population (that is, all employee jobs). Jobs that are not registered on Pay As You Earn (PAYE) schemes are not surveyed. These jobs are known to be different from the PAYE population in the sense that they typically have low levels of pay.

Consequently, ASHE estimates of average pay are likely to be biased upwards with respect to the actual average pay of the employee population. Non-response bias may also affect ASHE estimates. This may happen if the jobs for which respondents do not provide information are different from the jobs for which respondents do provide information. For ASHE, this is likely to be a downward bias on earnings estimates since non-response is known to affect high-paying occupations more than low-paying occupations.

Finally, ASHE results tables do not account for differences in the composition of different "slices" of the employee workforce. For example, figures for the public and private sectors include all jobs in those sectors and are not adjusted to account for differences in the age,

qualifications or seniority of the employees or the nature of their jobs, all factors that may affect how much employees earn.

***Returns***

Various procedures are in place to minimise errors in returned data. Returns undergo a range of checks that include validation against previous returns and expected values, selective editing (a technique for prioritising suspicious values for follow-up based on their impact on published results) and re-contacting businesses for verification. Similar checks are also made at the aggregate level for main results.

## Quality assurance processes at DCMS

The majority of quality assurance of the data underpinning the Earnings in the Digital Sector release takes place at ONS, through the processes described above. Once ONS have ensured all their in-house data checks, the data required by DCMS are sent via secure transfer. Further quality assurance checks are then carried out within DCMS.

Production of the analysis and 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***

As the ONS do not publish the raw data at each SIC code[4] level, the data is required to calculate the median annual pay for employees in the Digital sector. The SIC codes used to define the Digital Sector can be found in the DCMS methodology report[5]. DCMS discussed the 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 improved the understanding of the data at DCMS, helping us to ensure we receive the correct data and use it appropriately.

***Checking of the data delivery***

The data is delivered to DCMS as an SPSS file once the ONS have published their latest earnings release. For this particular release we check that:

- We have received all data at the 4 digit SIC code level, which is required for us to aggregate up to produce estimates for our Digital Sector and sub-sectors.
- Data at the 4 digit SIC code has not been rounded unexpectedly. This would cause rounding errors when aggregating up to produce estimates for our sectors and sub-sectors.
- Data for the correct year has been included
- The number of rows in the data seem sensible compared to previous year's data.

[4] https://onsdigital.github.io/dp-classification-tools/standard-industrial-classification/ONS_SIC_hierarchy_view.html

[5] https://www.gov.uk/government/publications/dcms-sectors-economic-estimates-methodology

***Data analysis***

At the analysis stage, data are aggregated up to produce information about the Digital Sector, Digital sub-sectors as well as the overall UK economy. The lead statistician builds in the following checks at this stage:

- Checking that the analysis for the overall UK economy matches that of the ONS's published outputs. This includes analysis by various demographics (such as age, gender, work region, gender pay gap)
- "Sense checks" of the data. E.g. does the median annual earnings in the Digital Sector and sub-sectors look similar to last year?
- Making sure it is not possible to derive sensitive data from the figures that will be published, especially at lower aggregations e.g. earnings by gender and by work region.

***Quality assurance of data analysis***

Once analysis is complete, the producer hands over to the quality assurers to carry out further checks of the analytical work completed. A detailed quality assurance (QA) log is produced by the lead statistician which documents the checks needed on the SPSS syntax, tables and report. Within the document, the suggested checks are listed and described, and there is space for the quality assurers to indicate the outcome of the checks and give any other feedback. After the publication, the quality assurance processes are reviewed to ensure they are relevant and comprehensive. The checks listed in the QA log cover:

- Ensuring the correct data have been used for the analysis e.g. has the 2018 SPSS data been used to derive the 2018 figures, or has the 2017 data been used by mistake?
- Checking that the correct SIC codes and SOC codes (used for occupations) are all accounted for and no codes are missing or included by accident.
- Checking the syntax picks up the correct variables e.g. checking the annual pay variable is used and not the weekly pay variable for analysis on gross annual earnings.
- Sense check of percentage change figures to the previous year – do they look sensible?
- Cross checking report figures with the published tables.
- Checking the charts are linking to the correct data.
- Making sure any statements made about the figures (e.g. regarding trends) are correct according to the analysis.

***Dissemination***

Finalised figures are disseminated as Excel tables and a written PDF report (which includes written text, graphs, tables and infographics) published on GOV.UK at https://www.gov.uk/government/collections/dcms-sectors-economic-estimates

## Next steps

We encourage our users to engage with us so that we can improve our statistics and the documentation surrounding them. If you would like to comment on this report on quality assurance processes, or have any enquiries please get in touch at evidence@culture.gov.uk.

Department for Digital, Culture, Media & Sport

**4th Floor**
100 Parliament Street
London
SW1A 2BQ