

# Quality assurance of administrative data (QAAD) report for Value Added Tax turnover data

Investigation of the administrative data sources used in the production of short-term economic output indicators by the Office for National Statistics' National Accounts and Economic Statistics Group.

Contact:

[gdp@ons.gov.uk](mailto:gdp@ons.gov.uk)

Release date:

11 April 2019

Next release:

To be announced

## Table of contents

1. [Introduction](#)
2. [Quality assurance of administrative data \(QAAD\) assessment](#)
3. [Areas of quality assurance of administrative data \(QAAD\)](#)
4. [Summary](#)

# 1 . Introduction

## 1.1 Background

The Office for National Statistics (ONS) uses a monthly delivery of Value Added Tax (VAT) turnover data from Her Majesty's Revenue and Customs (HMRC) to supplement Monthly Business Survey (MBS) and Monthly Survey of Construction Output data covering a wide range of industries. These data are used as a source of turnover in a range of economic statistics, including the calculation of the Index of Production, Index of Services, Index of Construction and the output measure of gross domestic product (GDP(O)) for the UK. We continue to investigate the use of VAT in other ONS economic statistics.

This report outlines the data process from initial collection through to the output of the release. It identifies potential risks in data quality and accuracy as well as details of how those risks are mitigated.

This report forms the latest in a series of quality assurance of administrative data (QAAD) reports produced by the ONS to communicate details of the administrative data sources that we use in the production of short-term economic output indicators. Such reports are a requirement set out by the [UK Statistics Authority](#). This report specifically focuses on our use of VAT turnover administrative data to supplement survey data in short-term economic output indicators.

Further information relating to quality and methodology for short-term indicators can be found in the appropriate Quality and Methodology Information (QMI) reports:

- [UK Index of Production QMI](#)
- [Index of Services QMI](#)
- [Construction output QMI](#)

Further information relating to quality and methodology for ONS uses of VAT can be found in our latest reports:

- [How ONS will use VAT](#)
- [VAT turnover data in National Accounts: background and methodology](#)

## 2 . Quality assurance of administrative data (QAAD) assessment

### 2.1 UK Statistics Authority QAAD toolkit

The assessment of our administrative data sources has been carried out in accordance with the [UK Statistics Authority's Quality Assurance of Administrative Data \(QAAD\) toolkit](#).

Each administrative data source investigated has been evaluated according to the toolkits risk and profile matrix (Table 1) reflecting the level of risk to data quality and the public interest profile of the statistics.

Table 1: UK Statistics Authority quality assurance of administrative data (QAAD) risk and profile matrix

Level of risk of quality concerns	Public interest profile		
	Lower	Medium	Higher
Low	Statistics of lower quality concern and lower public interest [A1]	Statistics of low quality concern and medium public interest [A1/A2]	Statistics of a low quality concern and higher public interest [A1/A2]
Medium	Statistics of medium quality concern and lower public interest [A1/A2]	Statistics of medium quality concern and medium public interest [A2]	Statistics of medium quality concern and higher public interest [A2/A3]
High	Statistics of higher quality concern and lower public interest [A1/A2/A3]	Statistics of higher quality concern and medium public interest [A3]	Statistics of higher quality concern and higher public interest [A3]

Source: Office for National Statistics

The toolkit outlines four specific areas for assurance and the rest of this report will focus on these areas in turn. These are:

- operational context and administrative data collection
- communication with data supply partners
- quality assurance principles, standards and checks applied by data suppliers
- producer's quality assurance investigations and documentation

In the assurance of our data source, we have chosen to give a separate risk and profile matrix score (Table 1) for each of the four areas of assurance. This will allow us to focus our investigatory efforts on areas of particular risk or interest to our users (Table 2).

## 2.2 Assessment and justification against the QAAD risk and profile matrix

Table 2: QAAD risk and profile matrix assessment of Value Added Tax administrative data used to measure short-term economic output

	<b>Low</b>	<b>Medium</b>	<b>High</b>
	<b>[A1]</b>	<b>[A2]</b>	<b>[A3]</b>
Operational context and administrative data collection			[A3]
Communication with data supply partners			[A3]
Quality assurance principles, standards and checks by data supplier			[A3]
Producers quality assurance investigations and documentation			[A3]

Source: Office for National Statistics

The risk of quality concern and public interest profile has been set as “high” because of the contribution that Value Added Tax (VAT) data make to the Index of Production (12.5%), Index of Services (6.3%), Index of Construction (14.3%) and gross domestic product (14.4%). As such, a score of A3 is deemed appropriate for this data source.

All scoring was carried out by the Office for National Statistics based on the level of risk of the data and interest of our users. This has used a balanced approach to capture both the quality dimension and public interest in the use of detailed administrative data in producing economic statistics. The data used follows a complex collection process with a large number of businesses providing the data, which is then re-purposed for input into calculating economic statistics. Results for each area of assurance for VAT turnover are shown in Table 2. If you feel that this report does not adequately provide this level of assurance or you have any other feedback, please contact us via email at [gdp@ons.gov.uk](mailto:gdp@ons.gov.uk) with your concerns.

## 3 . Areas of quality assurance of administrative data (QAAD)

### 3.1 Operational context and administrative data collection (QAAD matrix score A3)

This relates to the need for statistical producers to gain an understanding of the environment and processes in which the administrative data are being compiled and the factors that might increase the risks to the quality of the administrative data.

Her Majesty's Revenue and Customs (HMRC) is a non-ministerial department of the UK government that is responsible for the collection of taxes, including Value Added Tax (VAT). VAT was introduced in 1973 and is the third-largest source of government revenue. Businesses are required to register for VAT if their [annual VAT taxable turnover is more than £85,000](#) (since April 2017). As such, VAT is expected to capture data from businesses above the turnover threshold. In practice, there can be some exceptions based on the timing of the required returns. Additionally, businesses below the threshold can register voluntarily. HMRC receive VAT turnover data through section 91 of the [VAT Act 1994](#).

## HMRC Data collection

HMRC collect VAT returns using secure online digital forms, although some businesses that are unable to file online can request to submit VAT returns in other ways, for example, a paper filing. The ONS receives a file each month that contains VAT turnover data from all traders who have made a return to HMRC during the previous month.

VAT returns can be submitted for reference periods of either one month (approximately 10% of returns), three months (approximately 89% of returns) or a year (less than 1% of returns). Most businesses report VAT using a three-month reference period, starting from any month in the year. However, this is troublesome for short-term economic output indicators as we require a monthly output. These three-month VAT returns can refer to periods closing at the end March, June, September and December; January, April, July and October; or February, May, August and November. Of these three possible quarterly reporting patterns, the largest proportion, and in particular, large businesses (those with employment greater than or approximately equal to 150 for production, 250 for services or 100 for construction), report on the calendar end months March, June, September and December.

The ONS currently uses the monthly delivery of VAT data to improve the coverage of smaller and simpler, in terms of organisation complexity, businesses for selected industries. This delivery contains the raw uncleaned data as they appear on the VAT forms.

New methods have been developed by the ONS to use turnover data from the monthly delivery of VAT data to supplement data from 45,000 businesses that are selected as part of their Monthly Business Survey (MBS). Further explanation of the methods is given in Section 3.4. The increased number of returns from VAT means that more quarterly data are available at a granular level, for instance within some industries or size-bands. This will hopefully lead to a smaller MBS sample, which would reduce respondent burden.

## Online returns

Businesses log into their online accounts to submit VAT returns. There are [legal conditions](#) that apply to submitting the returns and many claimed benefits including:

- a confidential online account
- a safe and secure method of sending a return
- an on-screen acknowledgement that the return has been submitted
- an option to receive electronic reminders when your return is due
- automatic calculations to reduce errors when completing the return
- adjustments to VAT accounts to [correct errors](#) on past returns
- digital VAT records stored up to six years

A VAT form consists of nine boxes, which are shown in Table 3. For the purposes covered by this QAAD, we use data returned in box 6.

Table 3: Question boxes on the Value Added Tax return

<b>Box</b>	<b>Definition</b>
Box 1	VAT that businesses are required to pay on goods and services supplied within the period (within the UK)
Box 2	VAT that businesses are required to pay on acquisitions of goods from other EU members states
Box 3	Total output tax that businesses are required to pay (sum of boxes 1 and 2)
Box 4	Total input tax that businsses are entitled to claim for a period
Box 5	Net tax (the difference between boxes 3 and 4)
Box 6	Total outputs (sales) excluding VAT (including sales to other EU member states)
Box 7	Total inputs (purchases) excluding VAT (including purchases from other EU member state)
Box 8	Total value of all sales of goods to other EU member states
Box 9	Total value of all purchases of goods from other EU member states

Source: Office for National Statistics

The dataset provided to the ONS by HMRC each month contains the variables in Table 4. We use these variables to transform VAT turnover data to a level that is comparable with ONS surveys. This process is described in more detail in Section 3.4.

Table 4: Data provided to the Office for National Statistics by HM Revenue and Customs each month

<b>Variable</b>	<b>Description</b>
VAT period	Last month to which VAT return relates to
VAT reference	Nine - digit unique identifier for each VAT trader
Record type	Determines what is being returned on the VAT return
Stagger	Number between 0 and 15 detailing the month (s) in which HMRC expect returns for given VAT reference
VAT SIC	UK SIC 2007 that has been assigned to the unit by HMRC
Turnover	Total value (excluding VAT) of goods and services supplied in the period
Expenditure	Total value (excluding VAT) of goods and services purchased in the period
Receipt of date	Date on which return was received by HMRC

Source: Office for National Statistics

## Inter-Departmental Business Register data collection

The Inter-Departmental Business Register (IDBR) provides the main sampling frame for surveys of businesses carried out by the ONS and other government departments. The IDBR covers 2.7 million businesses in all sectors of the UK economy. The two main sources of input data are from the VAT and Pay As You Earn (PAYE) systems, as supplied by HMRC. All data on the IDBR are treated as Official Sensitive and are protected by the [Code of Practice for Statistics](#).

HMRC send the monthly file of VAT data on the first working day of each month. The file is transferred electronically via secure email to the ONS. The datasets that are received contain the raw (not cleaned) VAT turnover data as they appear on the VAT return forms. The VAT files are used to update and maintain the IDBR sample frame.

## Strengths

- VAT turnover is received by law through section 91 of the VAT act 1994.
- Digital data collection and VAT records are stored for up to six years.
- It is a safe and secure method for returns.
- The timeliness of quarterly returns, particularly for the larger companies.
- Businesses can correct errors on their VAT returns.
- There are monthly deliveries of data to the ONS.
- The number of records is greater than the Monthly Business Survey (MBS) sample-size for smaller businesses, potentially leading to more precise and timely statistics, whilst reducing the reporting burden on businesses.

## Weaknesses

- Business below the £85,000 threshold can register voluntarily.
- The data are raw uncleaned data as they appear on VAT return forms sent to the IDBR.

## 3.2 Communication with data supply partners (QAAD matrix score A3)

This relates to the need to maintain effective relationships with suppliers (through written agreements such as service level agreements or memoranda of understanding), which include change management processes and the consideration of statistical needs when changes are being made to relevant administrative systems.

### HMRC communication with registered VAT businesses in the UK

HMRC provide a multi-channelled approach for VAT account holders to provide and update their VAT returns, including how to correct any previous errors or make adjustments. General details about [how to fill in and submit a VAT return](#) are available.

### ONS communication with HMRC

The process of data transfer between the ONS and HMRC is via a monthly and a quarterly file containing the latest data. A memorandum of understanding between HMRC and the IDBR has been drawn up in line with the principles and protocols of the Code of Practice for Statistics. The purpose of the agreement is to provide input data that allow ONS to update and maintain our business sampling frame, and for use in the production of economic statistics and national accounts. Either party is free to propose amendments to the memorandum of understanding at any time, and it is subject to a full review annually.

Arrangements are in place that allow the ONS to query HMRC about data issues. HMRC offer a single point of contact for further investigation. Further contact arrangements are in place through email and phone, allowing HMRC and the ONS plenty of opportunities to communicate specific changes and issues.

## **Internal ONS communication**

The ONS National Accounts and Economic Statistics (NAES) team receives data from the ONS IDBR team, who are the initial recipients of VAT data within the ONS. The areas work closely together, including liaising with the ONS Data as a Service team as a central point of contact. The economic output areas have a long history of relationship building with their main stakeholders and users of the data. This is achieved by face-to-face meetings, email and phone calls.

## **Strengths**

- There is multi-channel communication between HMRC and VAT account holders.
- A memorandum of understanding between HMRC and the ONS has been drawn up in line with the principles and protocols of the Code of Practice for Statistics.
- There is a strong history of relationship building between HMRC, the IDBR and NAES.

## **3.3 Quality assurance principles, standards and checks by data supplier (QAAD matrix score A3)**

This relates to the validation checks and procedures undertaken by the data supplier, any process of audit of the operational system and any steps taken to determine the accuracy of the administrative data.

### **HMRC quality assurance checks on the quarterly VAT file**

HMRC carry out their own quality assurance checks on the VAT data on a regular basis. These are included within quarterly VAT file sent to the ONS in addition to the monthly files. However, this quarterly supply is too slow a supply for short-term indicators and the monthly file delivery is used instead. HMRC currently have no plans to quality assure the data received through this pathway.

For the quarterly delivery, a dedicated HMRC team is in place to handle the correction notices submitted by businesses. Each notification is subject to checks to protect the revenue and the interests of taxpayers generally. The HMRC team will contact businesses to clarify the details of the notices before they process them. In some cases, a visit may be necessary. HMRC officers take all steps to deal with any queries as swiftly as possible.

To overcome the above quality assurance issue with the monthly file, we carry out our own quality assurance checks using a new statistical platform and statistical methods developed by validation experts within ONS. These statistical methods build on the cleaning methods developed to clean and validate data from 45,000 businesses selected as part of their MBS (for further information see Section 3.4).

## **Strengths**

- The ONS has developed new methods to clean and validate VAT monthly turnover data.

## Weaknesses

- VAT turnover records held on HMRC systems will have been checked for completeness and accuracy, however, these checks are not sufficiently timely for the monthly delivery of VAT data that is used for the output discussed here.
- Limited knowledge and understanding of HMRC quality assurance with the quarterly VAT returns.
- The data delivered to the IDBR are raw and uncleaned and no further checks are applied.
- No quality assurance checks on the monthly file by HMRC for NAES purposes.

### 3.4 Producers quality assurance investigations and documentation (QAAD matrix score A3)

This relates to the quality assurance conducted by the statistical producer, including corroboration against other data sources.

When the ONS began using VAT data in the calculation of the output measure of gross domestic product (GDP (O)) in December 2017, turnover data were used from 630,000 businesses. This represented a significant advance in the transformation of UK National Accounts and short-term economic output indicators. In June 2017, VAT turnover data usage was increased to 750,000 businesses.

Following an internal review of our methodology and consultation with stakeholders, academic associates and international experts, the ONS agreed to combine output estimates from the MBS and Monthly Survey of Construction Output and the newly developed VAT turnover processing pipeline. The MBS will continue to be used for larger businesses, which also provides the ONS with the opportunity to gather business intelligence as part of the survey.

The NAES team within ONS receive raw (not clean) VAT turnover data from the monthly delivery of VAT data, from the IDBR team, with sufficient time being allowed to investigate and analyse the data in time for publication within the preliminary estimate of GDP(O).

New methods have been developed that allow ONS to use VAT turnover to supplement data from 45,000 businesses selected as part of the MBS. The processes applied to the VAT data are applicable to the data used within ONS for the purposes of producing economic statistics.

### New statistical processing platform

To process VAT data, a new statistical processing pipeline was developed, containing new methodologies. This pipeline transforms the monthly delivery of VAT data into a structure that is consistent with MBS. ONS are clear that the new VAT processing system is in the early stages of its development, and further investment is required to maximise the opportunities offered by the dataset.

We detail the different stages of processing the data in this section.

## Content editing

This cleaning of VAT data provides a processing challenge, in that the ONS cannot follow the same methods used for MBS data and contact the business to confirm their return. There is a trade-off between the greater coverage of the dataset and the infeasibility of the ONS being able to speak to businesses to gain anecdotal evidence and confirm erroneous data. As such, we have developed the following cleaning rules, and those described later in this section.

### Thousand pounds rule

Businesses should submit turnover values on their VAT return to HMRC in Great British pounds; however, it is possible for businesses to make clerical errors when submitting their VAT return by submitting in thousands of pounds. We use automatic methods to help identify these errors by dividing the current turnover value on their VAT return by the previous turnover value on their VAT return and calculating a ratio between the two. If the ratio is between 0.00065 and 0.00135, the turnover value for the current month fails the thousand pounds rule and is corrected by multiplying the value by 1,000. This rule is the same as that adopted by ONS business surveys in adjusting for thousand-pound errors on business surveys.

### Quarterly pattern rule

This rule identifies quarterly reporters not following a “true” quarterly pattern. There are three variations of this rule:

- a business reports the same turnover value in any three consecutive quarters followed by a different value in the fourth quarter (x,x,x,y)
- a business reports the same turnover value in any four consecutive quarters (x,x,x,x)
- a business reports a turnover value of zero in any three quarters followed by a positive value in the fourth quarter (0,0,0,x)

We change any turnover that follows one of these suspicious quarterly patterns by redistributing the turnover for that year in such that it follows the same seasonal pattern as other records from the same industry. The median turnover for businesses returning for the same quarter in the same industry are used for this correction.

## Matching and linking

VAT data are matched against a contemporaneous IDBR snapshot each month using a unique identifier – the VAT registration number of each business. This process enriches the source data with features from the IDBR that can be used in later processing of VAT data. Matching rates are over 99% of the VAT units in the VAT data each month.

## Apportionment

VAT returns can often be matched to multiple reporting IDBR Reporting Units. To apportion the VAT return to reporting units, frozen employment is used as an auxiliary variable and assumes a linear relationship between VAT turnover and employment.

## Calendarisation

Seasonal and trading days are used to calendarise the data. This method proportionally allocates quarterly and annual returns to months. The proportions are specific to each cell (aggregation of industry and employment size-band), and are derived from seasonal and trading day components estimated from cell-level MBS time series. VAT returns adopt a standard period through this approach, enabling a subsequent monthly estimation process.

The new statistical platform has transformed the dataset from an administrative dataset collected by HMRC, to a dataset that can now be directly compared with the ONS short-term turnover survey.

## Estimation

Ratio estimation using the IDBR frozen turnover variable is the estimation method selected for those cells that do not have a 100% response rate. This estimation method is already used within current ONS surveys by:

- calculating the amount of real turnover that has arrived in each cell, relative to the proportion of IDBR frozen turnover for those businesses that have provided a return in the cell
- the ratio is multiplied by turnover value for the cell to provide an estimate for real turnover missing from the cell for the target period

The estimation is assessed by applying the method as if it were a set point in the past and comparing the calculated estimate with the known “truth” when looking at the data in the present day.

## Cleaning and suspicious turnover checks conducted by the NAES team within ONS

After the data are processed using the new statistical platform, we use JDemetra+ software to seasonally adjust cell-level aggregates to identify outliers within the indices of production, services and construction. We compare VAT cell-level time series aggregates to the equivalent survey cells time series, and make a judgement call about which VAT time series need further cleaning.

We then run manual checks to identify anomalous reporters within industries. Using an internal database containing business correspondence and MBS data, we are able to make informed decisions in identifying businesses that have mistakenly included extra digits in their returns. We have made on average 50 manual adjustments to the data each month. This is minimal compared with the VAT selection, which is matched to 750,000 businesses.

We record an audit trail from when the data come in, and have regularly updated desk instructions in place that clearly explain the process. Finally, the final output is quality assured and signed off by two senior managers within the team.

## Cleaning and suspicious turnover checks conducted by the VAT team within ONS

Outputs from ONS processing of VAT turnover are used to produce value and volume estimates of output in different sectors of the economy. The data are delivered from the VAT team in NAES via the microdata processing system to the time series processing system. This incorporates the statistical processes of deflation and seasonal adjustment.

Once the data have been processed through the internal systems, the aggregate output is checked manually for growth rates and outliers, comparing with long-term and seasonal trends in previous years of VAT and survey data. This helps the team to gain a deeper understanding of the industries. They are also able to gain further anecdotal evidence of changes from the VAT team if needed.

The VAT team have clear desk instructions in place; they are regularly reviewed and updated. In addition to this, the final output is quality assured and signed off by two senior managers within the team.

## **Revisions Policy**

The revisions policy for short-term economic output indicators has been revised to minimise the frequency of revisions and ensure consistency with the accounts, as set out in the [National Accounts Revisions Policy](#). This ensures that the time series published for the production and construction industries will not receive two revisions in the same month, one being from the latest survey data and the second from the VAT turnover supplemented data used in the quarterly national accounts.

User engagement is continual, and the feedback tends to relate to the overall impact of the statistics rather than to the VAT data source used. To date, no specific feedback on the use of VAT has been provided.

## **VAT for regional economic indicator statistics**

As outlined in the Bean Review, there is also a need to produce more timely regional economic statistics and to make more use of administrative data sources in their compilation.

VAT turnover data will be used in an upcoming release of new experimental economic output statistics for the nine English regions, that is, Nomenclature of Territorial Units for Statistics (NUTS) level 1. The current aim is for VAT data to take precedence over Monthly Business Survey data. Work is being done to see how feasible this is for the cells where VAT turnover is not currently used in the National Accounts, with the intention of using VAT data from Quarter 1 (Jan to Mar) 2012 onwards.

The process of regionalising VAT turnover data follows on from the methods outlined in earlier in this section. This involves apportioning reporting unit-level VAT turnover to the geographical locations of economic activity (known as local units on the IDBR) that are covered by the reporting unit.

Here, local unit (LU) employment is used to apportion the turnover. The LU-level VAT turnover can then be aggregated by LU industry classification and LU region. From this, quarterly output indicator statistics can be produced for each English region.

The same local unit VAT turnover data are also used in the UK Regional Accounts measure of balanced gross value added (GVA(B)). Here, the extensive coverage of businesses allows the mapping of detailed industries down to lower-level geographic areas, corresponding to NUTS level 3 and UK local authority and council areas. VAT turnover data are used for all industries except the public services (public administration, education and health), households as employers, and imputed rental of owner-occupiers. The VAT data are used to map all years from 2011, with earlier years being modelled from these time series. The local authority-level estimates are used as building blocks to produce additional aggregates, providing estimates for combined authorities, city regions, local enterprise partnerships, growth deal areas and other emerging areas of economic interest.

## **Strengths**

- New methodologies to process VAT data have transformed the dataset from an administrative dataset collected by HMRC to a dataset that can be directly compared with the ONS short-term economic output surveys.
- Further checks are carried out by the ONS to ensure accuracy.
- The ONS checks previous VAT returns.
- A revisions policy is in place ensuring consistency with the UK National Accounts.
- There are clear and updated desk instructions.
- Final outputs are quality assured by two senior managers.
- There is continuous user engagement.

## **Weaknesses**

- Manual adjustments to the data can be made based on topic and conceptual expertise; however, the number of adjustments is minimal and the data would be unusable without these adjustments.

## **Next steps**

We will further develop our knowledge of the VAT dataset and investigate improvements to the current methodology.

# **4 . Summary**

The Office for National Statistics (ONS) considers the main strengths of the monthly delivery of Value Added Tax (VAT) data for use in short-term economic output indicators to be:

- VAT turnover is received by law through section 91 of the VAT act 1994
- there are monthly deliveries of data to the ONS
- Her Majesty's Revenue and Customs (HMRC) deliver data to the ONS through a secure online platform
- HMRC use digital communication with their customers
- a memorandum of understanding between HMRC and the ONS has been drawn up in line with the principles and protocols of the Code of Practice for Statistics
- there is a strong history of relationship building between HMRC and various ONS teams involved in the process
- the ONS have developed a new pipeline with new methodologies to process monthly deliveries of VAT turnover data, to produce an output dataset that can be directly compared with ONS short-term surveys
- the ONS carries out further cleaning checks to ensure quality
- there is a revisions policy in place ensuring consistencies with UK National Accounts

We believe that current limitations of this data source are:

- businesses below the £85,000 threshold are not required to register for VAT
- raw, unclean data are delivered to the ONS
- as part of the quality assurance process, the ONS can make several manual adjustments based on topic and conceptual expertise

In constantly seeking to improve our data sources, we will be taking next steps to investigate these limitations, and these will be communicated to users in future quality assurance of administrative data (QAAD) report updates for this topic.

However, based on the high risk of quality concerns and the high contribution that Value Added Tax feeds into the Index of Production (12.5%), Index of Services (6.3%), Index of Construction (14.3%) and gross domestic product (14.4%), ONS consider this data source to fulfil the requirements of an A3 assurance rating.