

Quality assurance of administrative data report for mining of coal and lignite and extraction of crude petroleum and natural gas industries

Report to investigate the administrative data sources we use in the production of short-term economic output indicators.

Contact:
Zuzka Hilton
stoi.development@ons.gov.uk

Release date:
4 March 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

Background

National Accounts and Economic Statistics (NAES) group within Office for National Statistics (ONS) receives data from the department of Business, Energy and Industrial Strategy (BEIS) on the mining of coal and lignite (UK Standard Industrial Classification (SIC) 05) and extraction of crude petroleum and natural gas (SIC 06) industries. These data form one source in the calculation of short-term economic output indicators, namely gross domestic product (GDP(O)) and Index of Production (IoP) for the UK.

This report outlines the process data take 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 NAES to investigate the administrative data sources we use in the production of short-term economic output indicators as set out by the [UK Statistics Authority \(UKSA\)](#). As such, this report focuses only on our administrative data used for the mining of coal and lignite (SIC 05) and extraction of crude petroleum and natural gas (SIC 06) industries. This report does not aim to cover other aspects of mining and quarrying collected by non-administrative data sources namely, mining of metal ores (SIC 07), other mining and quarrying (SIC 08) and mining support services activities (SIC 09), which are collected using the Monthly Business Survey and Work Force Jobs. Separate industries where we utilize administrative data will be considered in other QAAD reports in the series.

Further information relating to quality and methodology for the short-term economic output indicators is available:

[Gross domestic product, preliminary estimate](#)

[Index of Production QMI](#)

[Workforce Jobs Quality and Methodology Information Report](#)

Standard Industrial Classification (SIC) Overview

The mining of coal and lignite industry and extraction of crude petroleum and natural gas industry covers all activities under UK SIC 2007 division 05 and 06 respectively. Based on the “UK Standard Industrial Classification (2007)”, both of these industries are further classified to two sub-groups:

05 Mining of coal and lignite

- 05.1 - Mining of hard coal
- 05.2 - Mining of lignite

06 Extraction of crude petroleum and natural gas

- 06.1 - Extraction of crude petroleum
- 06.2 - Extraction of natural gas

According to the Inter Department Business Register (IDBR)¹ there were 10 enterprises classified under division SIC05, a decline from previous year when there were 15 Enterprises registered. All enterprises within division 05 were allocated to 05.1 - Mining of hard coal.

The IDBR has 125 enterprises classified under division 06. This is also a decline of 10 enterprises from the previous year. [The majority of enterprises within division 06 were allocated to 6.1 - Extraction of crude petroleum](#), which equates to 65 enterprises (55.5% of the total division).

Notes for: Introduction

1. The [Inter-Departmental Business Register \(IDBR\)](#) is a comprehensive list of UK businesses that is used by government for statistical purposes. It provides the main sampling frame for business surveys carried out by both ONS and other government departments. It is also an important data source for analyses of business activity

2 . Quality assurance of administrative data (QAAD) assessment

UK Statistics Authority QAAD toolkit

The assessment of our administrative data sources has been carried out in accordance with the [UK Statistics Authority 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).

Assessment and justification against the QAAD risk and profile matrix

Table 2: QAAD risk and profile matrix assessment of administrative data used to measure activities of unit and investment trusts

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

Source: Office for National Statistics

The risk of quality concern and public interest profile have both been set as “Medium” due to the contribution that mining of coal and extraction of petroleum (UK Standard Industrial Classification (SIC) 05) and extraction of crude petroleum and natural gas industries (SIC 06) feed into Index of Production (less than 0.01% for SIC 05 and 0.07% for SIC 06) and gross domestic product (0.9% for both SICs).

Volatility of data within the extraction of crude petroleum and natural gas industry has increased public interest. As such, a score of A2 (enhanced assurance) is deemed appropriate for these data sources.

All scoring was carried out by National Accounts and Economic Statistics (NAES) based on the level of risk of the data and interest of our users. Results for each area of assurance 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 [stoi.development@ons.gov.uk](mailto:development@ons.gov.uk) with your concerns.

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

Operational context and administrative data collection (QAAD matrix score: A2)

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.

National Accounts and Economic Statistics (NAES) use data delivered by the [Department for Business, Energy and Industrial Strategy](#) (BEIS) for the mining and quarrying industries. This ministerial department brings together responsibilities for business, industrial strategy, science, innovation, energy, and climate change and are supported in this by [46 agencies and public bodies](#). There are two main bodies responsible for operation and regulation in the production of statistics that are used for the mining and quarrying industries.

The Coal Authority (CA) is an executive non-departmental public body sponsored by BEIS. It owns, on behalf of the country, the majority of the coal in Britain. The Coal Authority are responsible for licensing coal mining in Britain, along with managing safety issues and dealing with water pollution caused by mining.

The Oil and Gas Authority (OGA) is an executive agency, created in 2015, whose role is to regulate, influence and promote the UK oil and gas industry to aid economic recovery for this sector. In October 2016 the OGA was incorporated as a government company with the Secretary of State for Business, Energy and Industrial Strategy the sole shareholder.

Mining of coal and lignite (SIC 05)

Monthly UK production statistics are produced solely from data obtained from the Coal Authority (CA). All coal producers with a licence have a legal obligation to declare their production to the CA, as stated in the [surface mining lease document \(PDF, 450KB\)](#). Coal producers cannot legally mine coal without a licence so must be registered with the CA; as such these data collected by CA are a census of coal production.

Coal Authority maintains a good working relationship with all suppliers. Reminder emails are sent to suppliers to make sure that data is ready to be despatched before deadlines. Data can be submitted electronically via email or manually when information is given over the phone. Data submitted undergoes various checks (for outliers, continuity and coherence) and can be queried at any point. Quality assurance checks by data supplies are described further in Section 3.

Coal Authority maintains a close connection with their suppliers. There is a helpline number that suppliers can call and an email address with staff trained to answer most queries. There is guidance within the lease document on what information is necessary, and there are also other documents available with further help and guidance. This industry is now reduced to a small number of suppliers and therefore a personal connection is easy to maintain.

Information collected by CA is automatically uploaded to a secured system. This system is shared between the CA and Department for Business Energy and Industrial Strategy (BEIS), and daily updates are sent automatically to the analysts in BEIS. There are several automatic validation checks that the coal data go through before they are used, which will be discussed in further detail in Section 3.

For each site, CA collect data on manpower (employees and contractors), production of coal, fireclay and other minerals (in tonnes), which is further split to “produced”, “disposals”, “to stock”, “ex stock” and “stock at month end”. They also collect information on transfers: “imported to site from” and “exported from the site to”.

Data collected by Coal Authority is used in several publications produced by BEIS including:

[Coal statistics](#)

They adhere to principals which underpin the [Code of Practice for National Statistics](#).

Data are delivered to NAES monthly via email and consists of raw data and an analysis of the industry with a commentary of important trends and findings. Updates for any data that have been previously forecasted are also included.

Strengths of Coal Authority (CA) data:

- legal obligation for suppliers to declare their production to the CA
- it is a census of all licence mines; it provides the full and comprehensive cover of the industry
- submitted data automatically uploaded to a secured system and daily updates are sent automatically to the analysts in BEIS
- follows [Code of Practice for National Statistics](#)

Limitation of CA data:

- they are manually checked and therefore open to the human error

Extraction of crude petroleum and natural gas (SIC 06)

Monthly UK production statistics are produced solely from data obtained from the Oil and Gas Authority (OGA). Monthly data for oil and gas production are collected through the automated Petroleum Production Reporting System (PPRS). PPRS is an administrative data collection system shared by OGA and BEIS that collects details of hydrocarbon (both oil and gas) production and is used to report flows and stocks of crude oils and natural gas liquids from field level through to final disposal.

It is a legal condition of operating on the UK Continental Shelf (UKCS) that companies must submit monthly returns to the OGA and as such these data collected by OGA are a census of oil and gas production. More [information about licensing](#) is available.

To encourage a high response rate within deadline for delivery to NAES there are automatic submission reminders within the PPRS; they are sent via email to the operator on the first day of the new submission month, then five days before the deadline and then every two days after the deadline if the return has not been submitted.

Information from the suppliers is sent via email (xml) and is then uploaded to the PPRS database by OGA specialists.

At the point of delivery to NAES (due to the required timeliness of the Index of Production (IoP) output) the coverage for this industry is around 80 per cent of the data, with strong forecasting mechanisms in place using other available sources. For gas production, BEIS use National Grid entry point data that is collected daily and illustrates how much gas is sent to the National Transmission System, except for a couple of terminals which are not connected via National Grid. This gives an accurate estimate, however, some estimation is needed if a terminal has UK and non-UK gas flowing through it, and consequently BEIS usually take a three-month average of the UK and non-UK ratio to calculate gas production via the terminal.

For oil production BEIS use a combination of the previous year's data, previous three-month average, the Treasury Economic Forecast, which includes field level production forecasts (only for oil production) and an element of judgement (for example, any industry intelligence such as a large field shutdown).

As soon as data is available, the forecast is replaced by real data.

Data that are being used in IoP are also published in several other BEIS outputs. These adhere to the Code of Practice for Statistics and are badged as National Statistics. These include:

[DUKES \(annual data\)](#) – Digest of UK Energy statistics. Provide a comprehensive picture of energy production and use over the last five years, with important series taken back to 1970.

[Energy Trends: gas](#). Energy trends is a quarterly bulletin containing statistics on all major aspects of energy in the UK.

Strengths of data:

- legal condition of operating on the UKCS that companies submit monthly returns to the Oil and Gas Authority
- data collected by OGA are a census of oil and gas production
- at the point of delivery there is at least 80% data available
- well established ways of estimating the missing data using other available sources

Limitations of data:

- data are currently manually uploaded to the BEIS Database and therefore open to human error

Communication with data supply partners (Risk QAAD matrix score A1)

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

Both Coal Authority (CA) and Oil and Gas Authority (OGA) have a good working relationship with data suppliers. They maintain regular contact and a high degree of engagement in different ways including:

- up-to-date contact information including a frequently asked questions page
- a published set of detailed user guides and guidance notes helping suppliers to understand and access their systems
- additional help and guidance in place and available on the website
- monthly contact through email and telephone to query data submission and data changes
- pre-booked face-to-face meetings to overcome barriers such as suppliers experiencing major difficulties providing data

The department for Business, Energy and Industrial Strategy (BEIS) have expressed they have a good working relationship with both the CA and OGA. They have established communication channels with face to face meetings, email communication and regular phone conversations. They also keep up to date records on specialists and important contacts in the respective Authorities. BEIS and OGA further collaborate as the system for data collection (the Petroleum Production Reporting System) is shared between the organisations.

There are also confirmed good two-way communications at each stage between Authorities and the suppliers, strengthened by the obligation of producers to deliver the production data in order for their operational licence to be valid. There is normally no need for BEIS to contact producer directly as the submission system allows for comments to be entered if data varies from previous returns.

NAES receives regular monthly updates with commentary and revisions from experts in BEIS detailing important trends and changes. Recent communication with the BEIS also resulted in them moving their deadlines for delivery of data to accommodate NAES's requirements for publishing of IoP.

Strengths of communication:

- two-way communication at each level
- operators have an obligation to deliver data for their licence to be valid
- submission system allows for comments to be entered
- there is a procedure in place to reduce respondent burden

Limitations of communication:

- NAES and BEIS are not in direct contact with suppliers, there are instances when BEIS can contact suppliers directly but the procedure has not been necessary due to the relationship OGA and CA have developed

Quality assurance principles, standards and checks by data supplier (Risk QAAD matrix score A1)

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.

Mining of coal and lignite (SIC 05)

Data submitted to the Coal Authority (CA) are manually entered to the database where they are checked by a specialist. They go through standard validation checks including but not restricted to: coherence, consistency, comparability, range checks and duplication. Data are then sent to the department for Business, Energy and Industrial Strategy (BEIS) where they are stored securely on the BEIS records management system. BEIS then conduct further validation checks on monthly and quarterly data before they are used to produce coal statistics, each return is checked both manually and automatically:

- to ensure the data balances out correctly, that is, supply equals demand
- for data anomalies, for example, there should be no negative figures, except for stock adjustments and opening stocks in a return are checked against closing stocks from the previous return
- against historical returns

Where there are discrepancies in the data these are investigated with the individual data suppliers. In circumstances where data providers are not able to provide an immediate answer, BEIS work with the data suppliers to provide the most accurate estimate until data issues are resolved. In addition to this, manual and automatic checks are carried out to also ensure all spreadsheets used in the production of coal statistic are linked correctly.

The Coal Authority dataset is relatively small and it is rigorously checked. Due to its current small size, procedures are relatively easy and quick.

Extraction of crude petroleum and natural gas (SIC06)

UK Continental Shelf (UKCS) operators submit their data electronically through the Petroleum Production Reporting System (PPRS), which has automated validation rules. If these are broken the system has two potential flags for the data, red and orange:

- if red then the system will not allow the return to be submitted until the data has been corrected
- If orange then the operator must provide comments to explain the reasons why the validation rule has failed

More often than not there will be no direct communication with the operator unless a discrepancy is discovered or a flag is triggered.

Once data are submitted to the PPRS system, the Oil and Gas Authority (OGA) perform sense checks on the data. These include checking if the density of oil is within expected parameters and comparing data with the previous month (if there has been a large change they will go back to the operator, however, the nature of production on the UKCS is extremely volatile so they take this into account.)

The department for Business, Energy and Industrial Strategy (BEIS) then take control of the rest of the quality assurance process and perform the following checks:

Check balance of input sheets for any given unit

Input sheets resemble the PPRS returns and returns should be a simple balanced breakdown for each associated gas terminal:

Associated Gas Entering Terminal	-	Gas Losses	-	Gas Flared	-	Gas Vented	-	Gas Utilised	-	Sales Gas to National Transmission System (NTS)	-	Natural Gas Liquids (NGL) Production	-	Sales Gas Non NTS
----------------------------------	---	------------	---	------------	---	------------	---	--------------	---	---	---	--------------------------------------	---	-------------------

The same applies for dry gas terminals, oil terminals and offshore loaders and so on. If this is not satisfied and the difference is relatively large, BEIS query the return directly with the reporting unit and ask for justification or corrected figures.

Comparison of data with “equivalent” external sources

For gas data within the PPRS return each month, reporting gas terminals will outline the amount of gas put into the National Transmission System; National Grid also independently collect and publish this information. In any given month, BEIS will compare the PPRS data with the “equivalent” National Grid data and compare any differences.

Gas balance

Each month BEIS create a gas balance where, in theory, calculated gas available should be equal to the amount of gas the pipeline operator’s record as being entered into the transmission system.

They calculate gas available from a combination of PPRS data and National Grid data (mostly PPRS) as gross gas production minus producer’s own use plus net imports. This should be approximately equal to the amount of gas input into the transmission system as independently provided to BEIS by the pipeline operators. Over the course of one year a difference of greater than plus or minus 0.5% is queried and they investigate “offending” months accordingly.

Oil balance and Natural Gas Liquids balance (NGL)

Similarly to the gas balance above, BEIS produce a monthly oil balance and natural gas liquids (NGL) balance, where they compare the supply according to PPRS data with the recorded demand as calculated from the independent Downstream Reporting System (DORS). For example, for crude oil, supply is equal to production plus imports minus exports plus stock change. This should be approximately equal to recorded demand provided by the refineries through DORS. Again, the tolerance for this across the year is plus or minus 0.5% and discrepancies greater than this are queried directly with the oil refiners and PPRS data providers.

Quality assurance from a technical user community and colleagues

The PPRS system is used to monitor field production by BEIS engineers and policy makers, and data from the system forms the backdrop to discussions on field by field production. Any anomalies will be spotted and commented on immediately as it is constantly monitored by BEIS and the technical user community.

Volatility of extraction of crude petroleum and natural gas data

Volatility is a particular concern in this industry and why a score of A2 has been given to this section of Index of Production (IoP). This is due to the potential for extreme volatility due to period on period changes of singular units that significantly influence the whole industry output and may impact the quality of the data.

For example, in December 2017, a closure of a major North Sea pipeline due to a hairline crack caused issues for over 80 platforms, resulting in production dropping by around 40% for the time when the pipeline was shut.

It is rare that one unit can make such a difference for the whole industry, however due to its nature and keen public interest, it is important to note this uncertainty in the data.

As this is a phenomenon of the industry profile and not the associated data sources, NAES analysts alongside BEIS and OGA specialists are always aware and informed about planned maintenance and can explain the drop in production.

Information about planned shutdowns and other known factors contributing to abnormal movement in data are included in a briefing that is sent to NAES with the figures.

Strengths of data:

- legal condition both for mining licence and operations on the UK Continental Shelf (UKCS) that companies submit monthly returns to the Authorities; therefore, it's a census
- automated checks on each stage
- regular audits by statistical teams
- previous years and rolling monthly period checks
- at the point of delivery there is at least 80% data available
- well established ways of estimating the missing data using other available sources

Limitations of data:

- in all stages data is currently manually uploaded to the BEIS Database, therefore open to human error

Producers quality assurance investigations and documentation (Risk QAAD matrix score A1)

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

Data for mining of coal and lignite industry (SIC05) and extraction of crude petroleum and natural gas (SIC06) is delivered to NAES within separate emails containing a CSV attachment by the assigned specialist for that industry from BEIS. Emails include data and commentary on any changes in the data from previous months, as well as any other important information, such as an explanation of zero figures if a plant is closed for maintenance.

Once the data is downloaded to Excel, NAES analysts carry out further sense checks to those previously conducted by Coal Authority, Oil and Gas Authority and the department for Business Energy and Industrial Strategy (BEIS) to make sure the data looks reasonable and in line with previous returns.

Once all checks are completed, data is uploaded into an internal Office for National Statistics (ONS) system as a CSV file, where seasonal adjustment is applied. There is no imputation to perform for this industry as BEIS provide us with their best estimates if data is missing. As a safeguard there is also an internal procedure of the final sign off before publication with the branch head. Data are then published within a mining and quarrying section in IoP. NAES keeps a full audit trail of any previous versions of the data and all previous CSV files are stored on an internal drive. Therefore, if there are any revisions, NAES are able to identify where and to what extent a change has taken place.

Strengths of data:

- sense checks – data are checked to ensure they are in line with previous trends
- full audit trail for both sets of the data

Limitations of data:

- occasional use of estimated data rather than actual values, offset by the requirement to deliver a timely data set and updated in following releases

4 . Summary

The National Accounts and Economic Statistics group (NAES) considers the main strengths of the data to be:

- high level of coverage for coal mining and oil and gas production
- high level of quality assurance by Coal Authority, Oil and Gas Authority and subsequently by the department for Business Energy and Industrial Strategy (BEIS)
- NAES quality checks and comparison with previous year's trends
- compliance with the Code of Practice for Official Statistics
- constant and productive cooperation and communication between NAES and BEIS, including changing deadline submission date

We believe that current limitations for this data sources are:

- only intermediate contact (via BEIS) of NAES with the Coal Authority and Oil and Gas Authorities who are collecting this data from suppliers
- manual entry of data

It must be acknowledged that results in this industry are highly volatile and therefore bring a lot attention. However, this uncertainty is an issue with the actual production in the industry rather than the source. Despite the other small limitations noted above, NAES considers this data source to be fit for purpose.