

Data Quality Management Policy



Data quality policy

Scope

This policy covers all the data that the Office for National Statistics (ONS) holds, either as a data controller or a data processor, including those obtained through surveys and from administrative sources in the private or public sectors. The policy is valid from data acquisition and collection by the ONS, and it also extends through the lifecycle up to and including data archiving and data disposal. This does not just include data sources we acquire or collect and the outputs we produce, but also any intermediate data product necessary to the data lifecycle. These data include data that are not collected or used for statistics and statistical research, such as HR and finance data.

This policy applies to all UK Statistics Authority employees, including ONS employees, which covers staff on fixed-term, temporary or permanent contract, staff on secondment, students, and contingent workers. Any user of ONS data (internal or external) must be encouraged to raise any issues related to data quality.

Background

This policy has been created to ensure that the quality of data is considered from the start, understood, assessed, communicated, and managed consistently throughout its full data lifecycle as highlighted by the data quality pillar within the ONS Data Strategy.

The [ONS Data Strategy](#) specifies our commitments for managing data, and the data policies support the implementation of these commitments. This data quality policy supports the delivery of the ONS Data Strategy, and specifically the implementation of the Data Quality Pillar, as well as setting the requirement for the management and governance of data quality in the ONS. This policy is in support of the [Code of Practice for Statistics](#), specifically its Quality pillar, which provides guidance to ensure that data and methods that produce assured statistics are adhered to.

It is important to ONS to establish and maintain a consistent approach to assuring data quality that covers the Data Management Association (DAMA) [Data Quality Dimensions](#). This is to support the overall strategic direction and achieve corporate objectives by using policies and data standards to uncover and address the issues and risks associated with the quality of the ONS data.

Finally, this policy covers the full data lifecycle and should be used in conjunction with the [Government Data Quality Framework](#).

Policy statement

Quality is a cornerstone of work at the ONS. We have a long history of ensuring quality in our statistical work, but trustworthy statistics must be built on data that are fit for purpose. We can and must do more to better understand the quality of all data we hold and use, and to demonstrate that we are a trustworthy organisation. We must work to better understand the quality of all data we use, collect, hold, and produce. The ONS must establish a culture that puts data quality at the core of every activity and remember that quality begins at the beginning of the data lifecycle. By ensuring that data quality is assessed at the earliest stages of planning for services, data storage, design, and data collection, we can enhance the trustworthiness of the data that we hold and use and build public confidence in our statistics.

Good quality data are fit for purpose: the data need to be good enough to support the outcomes they are being used for, and as stated in the [Government Data Quality Framework](#), “the level of quality required will vary depending on the purpose. Data quality is more than just data cleaning.”⁴

This policy demonstrates the need for data quality to be governed and communicated, and that data quality should be at the core of every activity.

Policy detail

1. Data quality shall be governed across the data lifecycle to ensure that data quality is appropriately managed:

- processes shall be in place for systematically reviewing and assessing the quality of data
- The impacts that quality has on decision making shall be considered, if the quality is low, it is essential to make sure that everyone is aware of the consequences of the risk
- data quality shall be considered and built into every activity; it must not be a consideration only at the end of the data lifecycle, for example, depending on data cleansing or editing
- any remediation activities shall be documented to include the impacts to the rest of the data lifecycle
- data quality issues shall be identified, documented, escalated and communicated; this is to include any third parties with which we share data

¹ The Government Data Quality Framework
<https://www.gov.uk/government/publications/the-government-data-quality-framework>

- processes shall be in place for reviewing the quality of data, this will be contained in the Data Quality Standard (owner to be confirmed)
- substantial changes to the Data Quality Management Policy shall be approved by the Data Governance Committee

2. Data quality shall be communicated to all users and data quality issues shall be documented at all stages of the data lifecycle, risks shall be known and understood:

- any changes or remediation of data quality shall be communicated to all users within the data lifecycle
- users shall be able to see the inputs used to create data products and shall have access to information on the quality of those sources
- all supporting data quality information shall be communicated, kept up to date, and its impact must be clearly explained to users
- supporting data quality information shall inform users, through tailored communication, to form their own opinions of the quality of the data, based on their specific needs
- all contracts and agreements for supply or use of data shall consider how to maintain, assess and communicate data quality, and escalate any issues
- the ONS must have in place robust and transparent procedures to capture feedback on data quality and resolve quality challenges by taking appropriate and proportionate actions
- a process that details the path to engagement and resolution shall be in place (process owner to be confirmed)

3. All teams in the ONS shall establish a culture of good data quality, meaning that data quality shall be at the core of every activity:

- data quality shall be part of all data management policies signifying its importance across the ONS
- every member of staff should be aware of policies relating to data quality and related policies, for example, data protection
- all ONS staff shall be empowered to challenge data providers where data is not fit for purpose; there shall be access to required data quality information so that users can make informed decisions
- [Data Quality Culture shall be considered at both ends of the data lifecycle](#); allowing information to be shared throughout the data lifecycle, and encouraging questions and challenges
-

Integrated Data Service

The above policy detail is also applicable to The Integrated Data Service.

The Integrated Data Service (IDS) will provide data across government for many uses, the quality of these data will be important in ensuring that outputs are of a minimum standard:

- data quality management shall be clearly defined and owned
- guidance shall be made available to users so that they are aware of potential escalation, should data quality problems occur
- users of IDS shall report data quality concerns so that they can be dealt with appropriately
- quality information and metadata shall be available at the same time as the data, regularly updated when necessary, and clearly communicated to users
- the Code of Practice for Statistics states that statistics should be based on data sources that are appropriate for the intended uses; inappropriate use of data shall be challenged
- the producer of the data shall do everything that they can to minimise the risk of misuse or misinterpretation

Roles and responsibilities

ONS Staff working with data

Staff working with data must:

- demonstrate compliance with the Data Quality Policy
- demonstrate awareness of their individual responsibilities contained within this policy and relating to data quality and quality assurance of data and their impact upon the quality of data, statistics, and analysis
- demonstrate awareness of their obligations under the [ONS Statistical Quality Improvement Strategy](#) to “ensure our data are of sufficient quality and communicate the quality implications to users”
- demonstrate awareness of the Quality Assurance of Administrative Data (QAAD) standard supporting the Code of Practice for Statistics principle on [Q1: Suitable data sources](#)
- raise concerns regarding data quality through line management and Quality Champions, and through the Statistical Quality Maturity Model Exercise

- ensure that risks and issues associated with data quality, together with associated actions, are appropriately recorded in their Divisional Quality Improvement Plan
- ensure that the commentary accompanying statistics and supporting [Quality and Methodology Information](#) appropriately reflect the quality of the data

Line managers

Line managers must:

- demonstrate awareness of this Data Quality Policy as well as other relevant policies and strategies in their specific areas
- monitor, challenge and assure that policy is being followed within their management
- support their staff in implementing processes that promote good data quality practices
- consider the quality of data when making or advising on decisions and ensuring that they have the information required
- collaborate across different business areas to ensure that data quality issues are addressed throughout the data lifecycle and not strictly within business areas
- maintain records of downstream users of any data products so that it is clear who data quality information needs to be shared with

[Government Data Quality Hub](#)

The Data Quality Hub is accountable to the Chief Data Officer and is responsible for:

- providing expert advice and guidance on what is best practice related to data quality
- acting as an ONS-wide sponsor for data quality, clearly communicating and coordinating data quality requirements
- supporting data quality remediation and escalation to the Quality Committee as and when required

Quality Committee

The Quality Committee is accountable to the National Statistician and the UK Statistics Authority (UKSA) Board and is responsible for:

- approval of complex data quality remediation as and when required
- mitigating tactical risks and assessing the mitigating actions for incidents related to data quality

UK Statistics Authority – Data Governance Legislation and Policy team

This team is accountable to the Data Governance Committee and is responsible for:

- providing independent scrutiny and assurance against the policy on behalf of the Data Governance Committee
- supporting data policy development to ensure that data quality is well represented
- providing independent scrutiny of data quality audits, as and when required
- ensuring that data quality activities remain transparent and align with this policy

Data Governance Committee

This committee is accountable to the National Statistician and UKSA Board and is responsible for:

- monitoring and reviewing how this policy is implemented, and acting on the assurance and maturity assessment information provided to them
- mitigating strategic risks and assessing the mitigating actions for incidents related to data quality

1. Appendix

2.1 Core data quality dimensions

Completeness

Completeness describes the degree to which records are present.

For a dataset to be complete, all records are included, and the most important data are present in those records. This means that the dataset contains all the records that it should and all essential values in a record are populated.

It is important not to confuse the completeness of data with accuracy. A complete dataset may have incorrect values in fields, making it less accurate.

Uniqueness

Uniqueness describes the degree to which there is no duplication in records. This means that the data contain only one record for each entity represented, and each value is stored once.

Some fields, such as National Insurance number, should be unique. Some data are less likely to be unique, for example geographical data, such as town of birth.

Consistency

Consistency describes the degree to which values in a dataset do not contradict other values representing the same entity. For example, a parent's date of birth should be before their child's.

Data are consistent if they do not contradict data in another dataset. For example, if the date of birth recorded for the same person in two different datasets is the same.

Timeliness

Timeliness describes the degree to which the data provide an accurate reflection of the period that they represent, and that the data, and data values, are up to date.

Some data, such as date of birth, may stay the same, whereas some, such as income, may not.

Data are timely if the time lag between collection and availability is appropriate for the intended use.

Validity

Validity describes the degree to which the data are in the range and format expected. For example, date of birth does not exceed the present day and is within a reasonable range.

Valid data are stored in a dataset in the appropriate format for that type of data. For example, a date of birth is stored in a date format rather than in plain text.

Accuracy

Accuracy describes the degree to which data match reality.

Bias in data may affect accuracy. When data are biased, it means the data do not represent the entire population. Data bias should be accounted for in measurements, if possible, and should be clearly communicated to users.

In a dataset, individual records can be measured for accuracy, or the whole dataset can be measured. This choice should depend on the purpose of the data and business needs.