

Quality Assurance of Administrative Data

England Tree Planting Programme

Daniel Braby/Kris Sales

Last updated: July 2024

Table of Contents

1	Introduction	2
2	Quality assurance of administrative data (QAAD) assessment.....	3
3	Areas of quality assurance of administrative data (QAAD).....	6
4	Summary.....	12
	Appendix.....	14

1 Introduction

1.1 Background

Forest Research (FR) receives data from the Forestry Commission (FC) and Forestry England (FE) on woodland creation in England. This report outlines the journey data undertake from initial collection through to the publication of the statistics. It identifies potential risks in data quality and accuracy as well as how those risks are mitigated.

The [UK Statistics Authority](#) have published a regulatory standard including a [Quality Assurance of Administrative Data \(QAAD\) toolkit](#). The standard was developed in response to concerns about the quality of administrative data, and in recognition of the increasing role that such data is playing in the production of official statistics.

Administrative data sources are combined in the production of official statistics on woodland creation in England. These are provided to us from FE and FC:

- FE's Geographic Information System (GIS) contains information on all woodland managed by FE. It is used to record, monitor, analyse and report on the extent and condition of the National Forest Estate.
- The FC is the main source for administrative data on woodland creation. FC currently collect data on woodland creation supported by government and other non-governmental organisations from 13 tree planting programmes in England.

Administrative data on new tree planting are combined with the [National Forest Inventory](#) (NFI) woodland map to determine total woodland area. Our quality assurance assessment is carried out on these administrative data sources only.

In this report the following terms are used:

- statistics producer refers to FR,
- data supplier refers to the FC or FE,

- data supply partners refers to those organisations that supply data to the FC.

2 Quality assurance of administrative data (QAAD) assessment

2.1 UK Statistics Authority QAAD toolkit

The assessment of the woodland creation administrative data sources has been carried out in accordance with the UK Statistics Authority QAAD toolkit.

Each administrative data source has been evaluated according to the toolkit's risk and profile matrix (Table 1), reflecting the level of risk to data quality and the public interest profile of the statistics. To determine which assurance level is appropriate for a statistics publication it is necessary to take a view of the level of risk of quality concerns and the public interest profile of the statistics.

Table 1: UK Statistics Authority QAAD risk and profile matrix

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

Source: Office for Statistics Regulation

The QAAD toolkit outlines four levels of assurance that may be required of a data set:

A0 – no assurance

The UK Statistics Authority states that the A0 level is not compliant with the [Code of Practice for Statistics](#).

A1 – basic assurance

The statistics producer has reviewed and published a summary of the administrative data QA arrangements.

A2 – enhanced assurance

The statistics producer has evaluated the administrative data QA arrangements and published a fuller description of the assurance.

A3 – comprehensive assurance

The statistics producer has investigated the administrative data QA arrangements, identified the results of an independent audit, and published detailed documentation about the assurance and audit.

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

- Operational context and administrative data collection,
- Communication with data suppliers,
- Quality assurance principles, standards and checks applied by data suppliers,
- Producer's quality assurance investigations and documentation.

Each of the four practice areas are evaluated separately, and the respective label of assurance is stated. This approach enabled an in-depth investigation of the areas of particular risk or interest to users.

2.2 Assessment and justification against the QAAD risk and profile matrix

The risk of quality concern and public interest profile of woodland creation statistics are rated by assessing: (a) the possibility of quality concerns arising in the administrative data that may affect the statistics' quality, and (b) the nature of the public interest served by the statistics.

The risk of quality concerns and public interest profile have both been set to "Medium" due to the number of administrative data sources brought together, differences in QA protocols, the regular coverage of woodland creation statistics in the media and impact on the [UK government's tree planting](#) statutory target.

All scoring was carried out by the Forest Information and Statistics team at FR.

Table 2: QAAD risk and profile matrix assessment of the administrative data used to measure woodland creation in England

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

Source: Forest Research

Therefore, as defined by the risk and profile matrix (Table 2), the combination of medium level of data risk concerns, and medium public interest profile indicate that enhanced assurance [A2] is the minimum level appropriate for woodland creation administrative data sources.

3 Areas of quality assurance of administrative data (QAAD)

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

This area 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 risk to the quality of the administrative data.

Forestry England data collection

Forest district staff at FE manually record new planting information using [Forester GIS](#), a FR tool for capturing and recording the extent and condition of woodland and trees. Forester GIS is a wholly digital service accessible on multiple devices concurrently. It facilitates the accurate and timely recording of new planting by foresters on-the-ground. Each new record must include information on area, planting year, and rotation number to be recorded successfully. New planting, by definition, refers to stands on their first rotation planted in the year of interest.

Data are aggregated for England and reported to FC using their submission form.

Forestry Commission data collection

The FC collect woodland creation data in three ways:

1. By querying records managed by teams within FC (Sustainable Forest Management, Finance or Evidence and Analysis).
2. By querying records managed by external organisations (i.e., Rural Payments Agency).
3. Return of a standardised submission form.

Regarding method (1), data on new planting under the English Woodland Grant Scheme, Woodland Carbon Fund, High Speed 2 Woodland Fund, and England

Woodland Creation Offer are collected from application registers and grant recording systems managed by FC.

Regarding method (2), data on new planting under the Countryside Stewardship scheme are collected by querying the Rural Payments Agency's Systems Data Warehouse.

Lastly, regarding method (3), data on new planting undertaken by FE, Environment Agency, National Forest Company, Northern Forest, Community Forests, and the Woodland Trust, etc., are collected by FC using a form submitted to data supply partners.

The accuracy and completeness of the data at this stage is crucial to the quality of the statistics.

A more detailed breakdown of FC data collection can be found in the [Appendix](#).

Strengths

- Forester GIS is a wholly digital tool making data collection efficient and minimising the chance of data entry errors. Information is verified independently either by other forest district staff or GIS technicians.
- A standard submission form ensures that all essential information is captured.
- Most data are collected quarterly, meaning data in FR's annual statistics publications have undergone four rounds of quality assurance prior to publication.
- Follows the [Code of Practice for Statistics](#).

Limitations

- Collecting data from any administrative system relies on the information entered being accurate and up to date.
- Coverage is not currently 100%. Several organisations/programmes do not have the infrastructure to report new planting.

- New planting on private land that is not supported by government (or third sector) funding is unknown.

3.2 Communication with data supply partners (QAAD matrix score A2)

This area relates to the need to maintain effective relationships with data 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.

FR and FE are both executive agencies of the FC. Agreement to share data for statistical purposes is covered in a Memorandum of Understanding on cross-border provision of forestry functions and research delivery between relevant ministers in Defra, FC, Scottish Government and Welsh Government (April 2019). Forest Research have expressed they have a good working relationship with both FC and FE. They have established channels of communication with regular meetings, email and phone communication.

In England, FC are the principal data suppliers and main channel of communication for woodland creation official statistics, responsible for collecting and supplying data to the Forest Information and Statistics team within Forest Research to agreed specifications. They liaise with data supply partners in respect to data collection, quality concerns and proposed reporting changes.

Strengths

- Two-way communication at each level.
- Data suppliers communicate regularly with data supply partners (at minimum quarterly).

- Data supply partners receive feedback during the data transformation process from statisticians, which increases the likelihood of issues being identified and corrected.

Limitations

- FR are not in direct contact with data supply partners.

3.3 Quality assurance principles, standards and checks by data supplier (QAAD matrix score A2)

This area 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.

Forestry England

An experienced GIS technician is responsible for collating and verifying new planting. Each quarter, a GIS technician reviews the administrative data, checking for data entry errors. To identify new planting, they filter the sub-compartment GIS layer for first rotation stands planting in the year of interest. They go through standard validation protocols, including but not restricted to, comparing the area of new planting with Ordnance Survey's MasterMap® and ESRI aerial photography, to ensure the geospatial vector data is accurately capturing the area of new planting and not buildings, roads, agricultural land, etc. If uncertainty remains, the GIS technician will contact the district planner for clarification.

The main risk to data quality is inaccurate recording of area, planting year or rotation number. Due to the small area of new planting each year by FE, the risk is considered minimal.

Forestry Commission

FC rely heavily on data supply partners to accurately record and report new planting. FC's Evidence and Analysis team use historical data, where available, to check the data match recent trends. Where both number of trees and area are

reported, they use one to check the other, ensuring that the number of trees planted is sensible given the reported area, and vice versa. Where the number of trees and area imply very high or very low spacing/stocking densities, the team will flag this with the relevant data supply partner.

In addition, FC now request geospatial vector data to compliment the completed form. This allows FC to verify new planting area against Ordnance Survey MasterMap® and ESRI aerial photography.

FC's outputs based on these data are also subject to review by the National Audit Office, which provides an added layer of Quality Assurance prior to checks by the statistical supplier.

Strength

- FE's sub-compartment layer is a 'living' record, undergoing constant review and revision.
- Data are checked by statisticians and analysts with an in-depth understanding of the forestry sector, meaning issues are identified and remedied efficiently.

Limitations

- Where only the number of trees is reported, area is estimated using an assumed stocking density, which varies between organisations/programmes and, in some cases, years.

3.4 Producer's quality assurance investigations and documentation (QAAD matrix score A2)

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

Pre-processing checks

The woodland creation administrative data are sent to FR within emails containing spreadsheet (xlsx) attachments by FC's Evidence and Analysis team. Emails include

data and commentary, as well as other important information, such as an explanation of zero figures and non-returns.

Data processing checks

The data are then copied into a master spreadsheet by FR statisticians, and further checks are made to ensure breakdowns (e.g., conifer/broadleaf) sum to totals. The data are checked independently by other FR statisticians and with the data suppliers (FC).

Output validation

As part of our output validation QA, the statistics are shared back with FC to ensure consistency between their statistical outputs and FR's.

Once all checks are completed, tabular representations of the new planting data are produced, broken down by public/private land and conifer/broadleaf. An additional level of quality assurance is afforded by producing the data visualisations for the statistics publication. Again, these outputs are checked independently by statisticians at both FR and FC.

As a final safeguard, prior to publication, the statistics are circulated among statisticians, analysts, and industry specialists with pre-release access for quality assurance (a [record of those individuals](#) with pre-release quality assurance access is available on the FR site). Data are then published in Provisional Woodland Statistics (in June) and Forestry Statistics (in September).

Strengths

- Multiple layers of quality assurance, both internally and externally.
- Full audit trail for the data, commentary and feedback.

Limitations

- Errors may be introduced at the data entry stage as this process is highly manual.

4 Summary

FR considers the main strengths of the woodland creation data to be that:

- Most data recording systems are well-established and frequently updated.
- The analytical teams at FR, the FC and FE carry out extensive quality assurance throughout the data collection and statistical production process.
- Methodological improvements are being introduced to allow corroboration with other data sources (e.g., Ordnance Survey, aerial photography).

The current limitations are that:

- There is limited understanding of what checks, if any, are carried out by some data supply partners.
- Woodland creation data are, in some cases, provided as number of trees planted. The area of woodland is estimated assuming a stocking density conversion factor.
- Areas of woodland less than 0.5 ha do not meet the NFI definition of woodland and are, therefore, excluded. There is potential for error here, either by failing to remove areas that are non-woodland or incorrectly removing areas that are woodland.
- FR's Forest Information and Statistics team are not involved in the data system design and have little to no direct engagement with the system designers.

In constantly seeking to improve woodland creation statistics, steps will be taken to mitigate the limitations identified in this report, and progress will be communicated to users in the next Quality Assurance of Administrative Data assessment.

However, woodland creation statistics are assessed as being assured to level A2 (enhanced assurance), as outlined in the Office for Statistics Regulation's QAAD toolkit.

If you are of the view that this report does not adequately provide this level of quality assurance, or you have any other feedback, please contact us via email at statistics@forestresearch.gov.uk with your concerns.

Appendix

Brief overview of current new planting programmes.

Government-support new planting programmes

Computed by FC's Evidence and Analysis team

[Rural Development Programme for England: English Woodland Grant Scheme \(2005 – 2014\)](#)

The Forestry Commission (FC) recorded information on woodland creation funded by the English Woodland Grant Scheme in their Grants and Licences Online System (GLOS) database. FC's Evidence and Analysis team employ pre-built queries in PowerBI to extract attribute information from the GLOS database, with payment date the key attribute for identifying new planting. The attribute data are then bound to spatial polygons to estimate area. To convert to number of trees, the Rural Payments Agency (RPA) assume a stocking density of 1,100 trees per gross ha. The database queries were built by an external company and rigorously tested. The data were updated fortnightly, and the likelihood of errors is considered low.

[Rural Development Programme for England: Countryside Stewardship \(2015 – 2020\)](#)

The RPA record new planting under the Country Stewardship scheme. The Evidence and Analysis team query the database for payments under Afforestation and Creation of Woodland, Woodland Tree Planting – Biodiversity, and Woodland Tree Planting – Improve water quality or reducing flood risk.

Number of trees is converted to area assuming a stocking density. This stocking density estimate is updated annually. No further QA is performed.

[Rural Development Programme for England: Countryside Stewardship, other new planting options \(2015 – 2020\)](#)

The Evidence and Analysis team also query the RPA database for payments under Planting Standard Hedgerow Tree, Planting Standard Parkland Tree and Woodland

Tree Planting to create or restore “landmark woodland.” Other new planting options refers to single trees.

Provided by FC’s Sustainable Forest Management (SFM) team England Woodland Creation Offer (2020 –Present)

FC’s SFM team report new planting under the England Woodland Creation Offer quarterly. New planting data are pulled from a spreadsheet (tracker) by filtering for the date the claim passed to the FC Finance team. The FC Finance team ensure that the new planting has been completed, and both number of trees and area are reported.

Woodland Carbon Fund

The SFM team record new planting under the Woodland Carbon Fund in their Woodland Carbon Fund Register spreadsheet. New planting is reported each quarter by filtering for “Number of trees” and all entries where the “Date Paid” column corresponds to the reporting year. Data are independently verified by a colleague.

High Speed 2 Woodland Fund

The SFM team collate new planting under the High Speed 2 (HS2) Woodland Fund in their HS2 Woodland Fund Register spreadsheet. New planting is reported each quarter by filtering for the “Number of trees in the claim” and all entries where the “Date Paid” column correspond to the reporting year. Both number of trees and area are provided. No further QA is performed.

Template submission to FC’s Evidence and Analysis team

For the remaining new planting schemes, the FC provide a form (Excel spreadsheet) to complete. The form requests: location, woodland type¹, woodland area (gross or net²) and/or number of trees, stocking density, tree species, and

¹ Woodland ([NFI definition](#)), small woods, groups of trees, single trees.

² Gross area: including any open space that is part of the planting design. Net area: excluding open space.

funding source of verified new planting in England that has been funded by the UK government and independently of that.

From September 2022, the FC have requested geospatial vector data (shapefile format), if available.

Environment Agency

The Environment Agency provide an aggregated total for new planting in England annually. In addition, geospatial point data are published on www.data.gov.uk, including latitude, longitude, area and/or number of trees, planting year and project. The Environment Agency undertake their own quality assurance. Stocking density is reported, allowing accurate conversion from number of trees to area and vice versa.

National Forest Company

The National Forest Company report new planting quarterly.

Northern Forest

Northern Forest report new planting quarterly.

Community Forests

The Mersey Community Forest is currently the only Community Forest scheme to report new planting in England. The form is returned annually.

Urban Tree Challenge Fund

The Urban Tree Challenge Fund report new planting annually.

Ministry of Defence

Currently unable to provide data.

National Highways (formerly Highways England)

Currently unable to provide data.

New tree planting support by other organisations

Woodland Trust

The Woodland Trust provide the Forestry Commission with an aggregated total for new planting each quarter. Data are aggregated and reported for England to avoid disclosure of private landowner information. Area of new planting is provided.

Woodland creation in England

Statistics are published in:

- Provisional Woodland Statistics (published annually in June).
- Forestry Statistics, Chapter 1: Woodland Area & Planting, along with the digest Forestry Facts and Figures (both published annually in September).
- England Tree Planting Program: Statutory Target Report (published every 5 years).
- FC's Key Performance Indicators (published quarterly).