

# Quality assurance of administrative data from Hubexo used in construction statistics

Quality assurance report produced by Economic Statistics Production and Analysis to investigate an administrative data source used in the production of short-term economic indicators.

Contact:  
Melanie Richard  
construction.statistics@ons.gov.  
uk  
+44 1633 456 344

Release date:  
13 January 2025

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](#)
5. [Cite this methodology](#)

# 1 . Introduction

## Background

Economic Statistics Production and Analysis (ESPA) within the Office for National Statistics collects data from [Hubexo](#) on new orders in the construction industry. This data will supplement our suite of administrative data sources of new orders and provide an additional source in the calculation of sub-national and sub-level type of work breakdowns for construction output.

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

This report forms part of a series of quality assurance of administrative data (QAAD) reports by ESPA to investigate administrative data sources we use in the production of short-term economic output indicators, as set out by the [Office for Statistics Regulation \(OSR\)](#). As such, this report specifically focuses on our administrative data for sub-level construction output data.

For further information relating to quality and methodology for the new orders in construction data, see our [New orders in construction Quality and Methodology Information \(QMI\)](#).

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

### Office for Statistics Regulation QAAD toolkit

The assessment of our administrative data has been carried out in accordance with the [Office for Statistics Regulation \(OSR\) Administrative Data Quality Assurance Toolkit](#).

The toolkit's risk and profile matrix must be used to evaluate each administrative data source that is investigated. The matrix allows us to assign a level of risk to data quality and the public interest profile of the statistics. For more information on the matrix, see the [Administrative Data Quality Assurance Toolkit Matrix](#).

The [UK Statistics Authority Administrative Data Quality Assurance Toolkit \(PDF, 243KB\)](#) outlines the assurance level framework for four specific data quality practice areas and the rest of this report will focus on these areas in turn. These areas are:

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

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

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

- Operational context and administrative data collection assessed as A1 (low).
- Communication with data supply partners assessed as A1 (low).
- Quality assurance principles, standards applied by data supplier assessed as A1 (low).
- Producer's quality assurance investigations and documentation assessed as A1 (low).

The risk of quality concern and public interest profile has been set as "low" because of the small contribution that these datasets contribute to the Office for National Statistics (ONS) construction statistics. Additionally, the sub national and sub sector dataset are no longer badged as a National Statistic, as shown in the [OSR's National Statistics Confirmation: ONS Construction statistics letter](#). As such, a score of A1 is deemed appropriate for this data source.

All scoring was carried out by ESPA based on the level of risk of the data and interest of our users. Results for each area of assurance for construction are shown in the previous list.

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

#### Operational context and administrative data collection (QAAD matrix score A1)

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.

Hubexo will be a chosen supplier of sub-national and sub-sector data in construction output to the Office for National Statistics (ONS). They are one of the leading providers of construction data. Hubexo have a strong history of information acquisition and relationship building with important customers throughout central government and the construction industry.

The service provided to Economic Statistics Production and Analysis (ESPA) enables our construction output team to have further coverage of UK construction projects and involved companies. Hubexo collect information on all new planning applications and decisions made across all Planning Authorities in the UK, as well as many separate unique and exclusive sources that are monitored daily, giving them a comprehensive picture of activity taking place currently, or in the future. Any future collection of new orders data for ESPA will be supplied through these existing and well-established processes.

To help their customers plan for the future [Hubexo also provides construction industry insights and market analysis](#), on a dashboard of indicators. For example:

- [The Hubexo Index](#)
- League Tables of top Contractors published in [Construction News](#) and [Building Magazine](#)
- the monthly [Hubexo Construction Review](#)

A fundamental part of Hubexo's business is to track construction projects and involved parties across the public and private sector.

Hubexo has a large, experienced team of over 150 web and telephone operatives, alongside a dedicated team scrutinising national, local and trade media. The business makes over one million telephone calls and conducts around 250,000 hours of research every year. This is to locate and engage key decision makers on projects to gather their database of every planning application and thousands of non-planning projects across the UK. They use this data to deliver insight and up-to-date construction leads to their customers. The value of the new orders data is that it measures main contractors by type of work and region.

The main users of the data are:

- those working directly in the construction industry who require details of current and future activity
- industry analysis requiring a leading indicator of the UK construction industry
- trade associations making UK and international comparisons
- other government departments such as [Department for Business and Trade](#), and the [Department for Levelling up, Housing and Communities \(DLUHC\)](#)

Hubexo are a member of the Office for National Statistics Construction Statistics Steering Group and have partnerships with the Building Research Establishment and the Considerate Constructors Scheme.

## Hubexo data collection for planned projects

The primary source for initially identifying construction projects is the planning process. Hubexo's data mining tools use Machine Learning and Artificial Intelligence. This is used to scrutinise thousands of planning documents, searching for key words in context and applying named entity recognition, alongside other techniques to extract useable content from this unstructured data. It is notable that at this stage, contracts are not awarded, and further investigation is needed through further telephone work. The data is collated daily, and values are taken at current prices (for example, no impact of price changes are accounted for).

Once the data is collated, Hubexo also operates an exhaustive process daily to review the past six months for applications which were submitted, but for various reasons, did not get posted at the appropriate time on the Local Planning Authorities (LPA) website. The publication of applications may be delayed because of validation checks by the LPA, or issues with payment of planning fees. Any amendments to existing planning applications will be allocated a new project identifier, if the existing project lifespan has ended. In addition, every project for 10 or more residential units is reviewed by a team of data and phone researchers.

In the case of major projects, Hubexo creates an initial project record for the full project. Once it becomes clear that phases or separate contracts are to be awarded at different times for different elements, further "child projects" will be created and the overall scheme marked as a "Masterlead" or parent project. The value of the contract awards will only be taken from the "child projects" to ensure no double-counting takes place and that the appropriate contract values are apportioned to the appropriate month.

Hubexo's operations, in both the UK and in several global data centres, act as data processor sites and are responsible for the initial data filtration, following the extraction of data from planning documents. Its UK-based research teams are led by sector experts (for example, housing, public sector, offices and so on) who oversee the work produced by their respective teams. Output from the research team is scrutinised by quality control teams, and random sampling is used to select a proportion of all projects to be reviewed and edited by the sector experts prior to publication.

Hubexo's research team use a robust recall system to track every large project through to either confirmed completion, or to a stage at which it is confirmed the scheme has been abandoned. A large project is considered to be a project with 10 or more residential units, or a build cost for non-residential schemes of £250k or more. All smaller projects are identified and tracked through the planning process to decision stage. A small project is considered to be a project with below 10 residential units or £250k build cost for non-residential schemes.

All larger projects (10 or more residential units or £250k or more build cost, for non-residential schemes) are fully telephone researched and tracked through from the earliest stage (often pre-planning) through to completion. This process involves direct telephone research contacting the key decision makers (such as architects, project managers, surveyors and so on) on each project, on a regular basis every few weeks to determine the latest position on projects. Once collected, the database is updated so an accurate representation of the progress of each project is maintained.

## Hubexo data collection for projects outside of the planning process

Telephone operatives make over 1 million calls per year, contributing to approximately 250,000 hours of research, engaging identified contacts from planning documents about the work they are actively involved in. Hubexo takes advantage of its existing relationships with contractors to get information on work they are bidding for. Businesses share information, and in exchange, this ensures comprehensive coverage on competitors and projects out for tender. Hubexo also makes use of a suite of additional and exclusive data sources to capture changes to existing projects and any projects that may have bypassed planning for several reasons.

## Hubexo project creation and tracking

Once projects have been identified, they are categorised into:

- schemes worth more or less than £250,000 for non-residential projects
- schemes with less than or 10 or more residential units

All projects are coded and categorised to identify different types of work taking place. Hubexo has mapped its in-house categorisation to the ONS taxonomy for 23 type-of-work examples, which include private and public-sector housing and infrastructure such as water, sewerage, and electricity. For more information on example of types of work, see Section 7 of our [New orders in construction Quality and Methodology Information \(QMI\)](#).

All planning applications of 10 or more residential units or units worth £250,000 or more, are passed to Hubexo's in-house research team. This team consists of web and telephone operatives that track the contract until completion or to a stage of confirmed abandonment. The team is focused on gathering important project data to provide their customers, including the ONS, with as complete an understanding, as possible, of activity taking place. This includes:

- contact details for relevant parties (name, job title, email, telephone, and mobile number)
- contractor awarded main contract
- timing details (start, duration, finish, and tender return dates)
- location
- consultant information
- material detail
- value of the main contract

These are collected for all projects with particular emphasis on project value, categorisation funding, and location for the purpose of ESPA requirements.

To ensure fully comprehensive coverage, information is also gathered from sources other than planning, for example:

### The Contracts League

The Contracts League, curated by Hubexo, was published in [Building magazine](#), [Construction News](#), [Housebuilder](#), [Architects Journal](#), and so on. This feature details those contractors, architects or clients who have won or awarded the most work each month. In compiling these features hundreds of businesses send Hubexo details of all projects won to ensure they feature most accurately in the final League tables.

### The Official Journal of the European Union

The [Official Journal of the European Union](#) or similar "Find A Tender" UK sites, which includes details of public sector tenders and new orders as they are placed.

## A suite of data sources exclusive to Hubexo

A suite of data sources exclusive to Hubexo, for example the [Considerate Constructors Scheme](#). This includes the best practice code of conduct, which the construction industry adopts. Contractors register their sites with the Considerate Constructors Scheme prior to commencement (sometimes months in advance), and the information on the project: client, contractor, start and end dates, and size are all recorded directly from the contractor, so the information is 100% accurate. Data, regarding approximately 7,000 projects a year, is shared exclusively with Hubexo. It is used to ensure projects are updated to the appropriate stage, but it also identifies hundreds of projects every year, which are either prior to planning or may bypass planning for various reasons.

Additional research is carried out to identify future planned projects, funded by:

- local authorities
- government departments
- deregulated industry
- the largest property owners in the private sector

Getting this information early helps Hubexo to gain valuable insight into up-and-coming new work.

The approach described in this section allows Hubexo to collect information on every planned project application. However, for projects outside of the planning process there are no guarantees that some projects will not be missed. However, every effort is made to try to include projects that bypass the planning process via the data sources named previously.

Hubexo estimates that the market make-up is split, approximately, 80% on planned projects and 20% non-planning. Approximately, half of all projects are identified from planning. The remaining half are identified by the other sources outlined previously, many of which will eventually go through planning, but many will bypass planning completely (conversions, refurbishments and so on).

Hubexo deliver data on a quarterly basis including revisions on the previous two quarters, which allow ESPA to produce timely outputs, with more reliable data from the next publication. One of the benefits of the sub-level construction data is that the data sources used in the model have contracts which are tracked from order to completion, so that any cancellation to contracts are reflected in revised data in future outputs.

The strengths of this method are that:

- information is collected on every planned project
- data mining tools are linked to all local authorities
- the research team collates data from contractors, and conducts further follow up validation and research
- data collected from other data sources (non-planning process) are collated and provide additional quality control
- there will be additional research for future planned projects
- all projects are coded and categorised to ONS taxonomy for 23 types of work
- timely quarterly deliveries show contracts tracked from order to completion

The limitations of this method are that:

- an unknown proportion of non-planning projects are potentially being missed
- there are data which are received from contractors on a voluntary basis
- smaller projects are also not tracked through to completion, so there may be some downward bias in valuation where costs are often upscaled through a project's lifecycle

## **Communication with data supply partners (QAAD matrix score A1)**

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

## **Hubexo communication with data providers**

Hubexo have developed and maintained close relationships with contractors, and other information providers, to secure exclusive access to data sources. Access is bound by service level agreements with providers. These ensure comprehensive coverage of all projects in the construction industry.

Hubexo upholds quality of communication with data providers, with a team of dedicated staff listening-in to call recordings made and marking them on a variety of measures around call quality, questioning techniques, and the comprehensiveness and accuracy of the information captured and input to the database.

All Hubexo research staff go through an intensive initial 13-week introduction training course. This covers three key areas:

- telephone researching - calling practices, questioning techniques and so on
- construction industry - the UK planning process, procurement routes, roles played by each company, technical terms, building materials, types of funding, and so on
- Hubexo's content management system - how to input or update data, searching techniques, quality control processes and so on

Following this the research staff begin on a simple project, such as updating tasks, and gradually rise through a structured series of levels, with defined quality gateways, as they gain more experience and show expertise.

## ESPA communication with Hubexo

ESPA in the procurement and research of the data have been happy with the answers and communication with Hubexo. During the data supply, ESPA can query certain projects or trends, to which Hubexo responds quickly with clarification and sometimes correction, if the subsequent quality assurance identified an issue.

The ONS Construction output branch has a good understanding of the processes carried out by Hubexo because of the information they have provided in meetings to define a working relationship. This is accompanied by routine check-ins to discuss and query information that is shared. ESPA are also able to attend the office for in-depth discussions if needed. If there was a change to the methodology, ESPA would expect to be notified by Hubexo, especially if there were any implications for the data.

Hubexo have agreed to produce detailed management information covering data quality and analysis, explaining significant movements in the data. Should the ONS identify any data as inaccurate, then Hubexo guarantees to re-research the detail and correct as required. As an assurance of the efficacy of this service, Hubexo offers a "Hotline" service to all clients who wish to query data or request an update, and this currently generates around 50 requests per working day.

Current average turnaround on these requests is 0.8 of a day, with around 75% being responded to within the day. Any queries from the ONS (as a key client) would be prioritised and responded to immediately within Hubexo's research team. Any inaccurate data identified would result in a review of processes to identify remedial action required, extra training or new exception reporting. This procedure would be recorded and discussed at the next regular meeting, to share what actions have been taken, to avoid a repeat occurrence.

A legal contract will come into effect detailing ESPA statement of requirements. Hubexo is contractually obliged to supply initial new orders estimated by the required period, within an agreed timeframe. They are required to supply ESPA with detailed management information, covering data quality and analysis, explaining significant movements in the data.

The strengths of this method are:

- strong communication links with data suppliers, backed by service level agreements
- regular telephone calls with data suppliers
- performance standards for telephone operatives, that ensure consistency and accuracy
- that the ESPA has a good relationship with Hubexo, with visits to both offices possible, if conditions permit
- that systems are in place for ESPA to query data



## Quality assurance principles, standards, and checks by data supplier (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.

All data are recorded through a secure database system. Sector experts have oversight on a range of key performance indicators across every construction project that is captured from data collation. These indicators are used to ensure the most important information of projects is sought first and prioritised when tracking a contract's lifespan. This includes tenders, timescales, cost, funding, and materials. It also includes contractors, subcontractors, and specified contract leads.

To supply new orders data, which are the input data for the sub-level construction output estimates, records are extracted and all projects worth more than £40 million are fully re-researched. An additional, random 10% of all projects are re-researched each quarter, to ensure data accuracy and integrity. Validation checks are carried out over all projects to cover mandatory information and ensure that fields are correlated to ensure consistency and accuracy. Examples of these checks include cross-referencing between value and floor area or number of units, and checking that those marked as "contract awarded" have a contractor listed and so on.

Quality assurance reports run in real time to spot any issues and enforce procedures. Any issues identified are re-researched, and the database is updated daily with a check or report created to ensure such risk is mitigated in the future.

All projects have a full "project history" recorded against them, which contains details of every field changed, what it was changed from and to, when the change was made and by which researcher or process. This allows Hubexo to review any issues arising and discuss remedial works.

Hubexo has a robust and tested Business Continuity Plan (BCP), should any external issue threaten its ability to continue to function. This BCP is tested on a regular basis, several times each year, to ensure the infrastructure is sufficiently robust to enable all staff to work from home if required (as was operated because of the coronavirus (COVID-19) pandemic). Because of the global geographical distribution of Hubexo's operations, any issues are unlikely to impact all data centres simultaneously. Hubexo's processes are set up to run 24 hours a day, seven days a week. This includes holidays, utilising the time zones of offshore operations to minimise any delays.

The strengths of this method are that:

- comprehensive quality assurance checks are in place
- larger and major projects are fully researched
- there is a reduction of delays in reporting

However, Hubexo pays less attention to smaller projects because of:

- volume
- the lack of demand for information on these projects
- the impact on timeliness of outputs they produce

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

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

On receiving the data from Hubexo, ESPA transfers the data into a pre-prepared spreadsheet, where a macro file is run to convert the data into a format that can be run on our internal software system.

Once the data has been processed through the internal software, the output is checked for growth rates and main trends. This is then checked against the original spreadsheet to ensure the new output has not changed because of the processing.

The data is supplied with detailed briefing information from Hubexo; using this and any trends picked up in the data, further questions are usually sent to Hubexo. They are quick to respond to queries, some of which result in further quality assurance work at their end. If the checks result in any revisions to the data, the data will be re-processed in the internal system.

The results from this process are then exported into another internal database where processing occurs in a table in the internal database's software. This is then checked against a separate spreadsheet that mimics the steps, to ensure the internal processes have run consistently.

The [Output in the construction industry: sub-national and sub-sector dataset](#) is a quarterly release within the monthly [Construction output in Great Britain bulletin](#). Once the publication is ready it is checked by the construction team and sent for comments to senior members of staff. Prior to publication, the release is uploaded to a secure preview site, where the data can be further checked to ensure it is correct.

ESPA can explain changes to the data because of the briefing provided by Hubexo, along with the delivery of the data, which includes details of all projects. Although these are not published for disclosure purposes, this helps ESPA gain a deeper understanding of the industry. They are also able to gain further anecdotal evidence of changes from Hubexo if needed.

ESPA have clear desk instructions in place, they are regularly reviewed and updated, and in recent months have undergone a full overview from task initiation to publication of data. In addition to this, Hubexo send the data file to at least two members of the team each month. Hubexo are also made aware if a specific member of staff should be sent the data to ensure receipt.

The strengths of this method are that:

- there are clear desk instructions
- there are comprehensive quality assurance checks in place, including handling revisions
- Hubexo provides a detailed analysis
- data is resupplied if errors are found
- final checks are undertaken by sector experts

## 4 . Summary

In investigating this administrative source for sub-level construction output data, Economic Statistics Production and Analysis (ESPA) considers the main strengths of the data, for its purposes, to be that:

- there is detailed knowledge and research of the subject area by Hubexo
- data is tracked from planning to completion
- there is coverage of every planned project application
- comprehensive quality assurance checks are carried out by Hubexo and the ESPA
- there is regular contact and briefings between construction suppliers, Hubexo and ESPA
- detailed analysis is provided by Hubexo
- there is a contract in place, detailing ESPA requirements

We believe that the current limitations of this data source are that:

- possible unplanned projects are being missed
- there is less attention for smaller projects, because of the timeliness pressure of outputs

We will continue to investigate these limitations as best practice, communicating relevant changes to users in future quality assurance of administrative data (QAAD) report updates for this topic.

ESPA considers this data source to fulfil the requirements of an A1 assurance testing, as these slight limitations are offset by the low risk of quality concerns and the small contribution in terms of sub-level construction output data in terms of the ONS Construction Statistics overall.

If you feel that this report does not adequately provide this level of assurance or have any other feedback, please contact [construction.statistics@ons.gov.uk](mailto:construction.statistics@ons.gov.uk).

## 5 . Cite this methodology

Office for National Statistics (ONS), released 25 July 2023, ONS website, methodology, [Quality assurance of administrative data from Hubexo used in construction statistics](#)