Qflow

THE STATE OF DATA QUALITY IN CONSTRUCTION

Construction Info Series
January 2025

GARBAGE IN GARBAGE OUT

Why data quality matters and why should you care?

Poor-quality data leads to flawed outcomes. In today's data-driven world, decisions across industries rely heavily on accurate, reliable data. Poor-quality data—whether outdated, incomplete, or full of errors—undermines decision making, leading to costly mistakes and inefficiencies. Caring about data quality isn't optional; it's foundational. By prioritising accurate, complete, timely, consistent and valuable data organisations and individuals can actually manage processes and decisions. "You can't manage what you don't measure" and in the end, the output is only as good as the input. This report looks deep into the state of data quality in our industry and offers vital insights on how we can all do a better delivering a more efficient and sustainable future for construction.

Foreword

In today's digital age, data has become the lifeblood of industries worldwide, and the built environment is no exception.

Our industry is embracing digital tech for data capture, auditing, and navigating new reporting demands, but data **quality** is often forgotten about in the race for **quantity**. With Artificial Intelligence promising exciting opportunities for all industries to do more with less and quicker, the need for high quality data has never been greater! Without it we are at risk of 'garbage in, garbage out'.

The built environment shapes how we live, work, and thrive, forming the backbone of communities and economies. With global building stock expected to double by 2060 (Hageneder 2020), and existing buildings and infrastructure in desperate need of upgrades, the construction sector must tackle mounting environmental impacts, economic pressures, and professional responsibilities head-on.

This is a sector entering a new era. An era characterised by stricter safety, sustainability reporting, net-zero goals, digital innovation, alongside the ongoing need for efficiency and profitability.

Navigating new technologies and responsibilities starts with reliable data. In construction, poor data can cause inefficiencies, safety risks, and setbacks in sustainability—challenges that ripple across projects and the industry's ability to meet its goals.

This report sheds light on the state of data **quality** in the construction sector, focusing on material deliveries and waste removals. Materials constitute 40% of a project's budget and account for 90% of the industry's embodied carbon emissions; it is a huge, complex and often opaque part of the industry that is often categorised as 'too hard, don't know how to fix'. Better quality data can help us dramatically cut carbon and save time and money while delivering projects.

By evaluating six years of data including over 1 million data points, against seven principles for data quality, the report provides a comprehensive overview of the current data quality landscape in construction.

By highlighting the critical importance of data quality and offering actionable solutions that digital technologies can unlock, we aim to inspire a sector-wide commitment to improving data quality.

This shift is essential for an industry that is able to leverage the potential of AI and drive towards a more sustainable, safe, and efficient future.

Brittany Harris,
CEO and Co-founder at Qualis Flow

EXECUTIVE SUMMARY

Today's world is driven by data, with global data production doubling every two years (Forbes, 2019). The Architecture, Engineering, and Construction (AEC) sector is no different. Digitalisation and the adoption of new technologies have turned the AEC sector into one which creates and collects massive amounts of data. The industry is now starting to use data to improve reporting and profitability and support better decision-making. However, this is only possible with high-quality data. **This report explores the state of data quality in the AEC sector, its implications, and actionable solutions to improve.**

Data quality issues in construction can lead to serious consequences. In fact, 1 in every 3 poor decisions in construction is a result of bad data (Autodesk, 2020). This is not just a theoretical problem; low-quality data can lead to costly inefficiencies. A staggering **$1.84 trillion was lost** globally in 2020 due to bad data in the construction industry alone (Autodesk, 2020). With materials accounting for over 40% of a construction project's capital expenditure (TrackUnit 2024), poor management of materials, enabled by poor data, results in 5%-11% budget wastage (Business Insider, 2017).

Beyond the financial implications, poor data, especially for construction materials, also puts safety at risk. The stakes are incredibly high when it comes to data quality for materials in construction. Without a consistent thread of information on all aspects of a building project, **safety considerations can be missed,** and accountability around who is responsible for enforcing these can be hard to pin down. Real-world disasters like the Grenfell Tower fire illustrate the **devastating consequences** when information about construction materials is not available to everyone who needs it.

The failure to track and report data on delivered materials also impacts sustainability efforts and adherence to industry targets around net zero. As deadlines around lowering emissions approach, improving data quality for carbon reporting will be vital.

To investigate data quality issues in construction, this report has evaluated 831,206 product deliveries and 195,178 waste removals across multiple construction projects. The data was assessed against seven key principles of good data quality: accuracy, completeness, uniqueness, consistency, validity, availability, and value. These principles serve as a foundation for understanding what constitutes "good" data in the construction industry and how it can be improved.

Due to the significant implications of data quality issues for construction material data, the findings resulting from the analysis of the product and waste documents are used as an example to showcase the challenge in the industry at large.

Our analysis uncovered some alarming trends. Of the deliveries reviewed, 95% contained “bad” documentation, meaning the data was incomplete, inconsistent, or inaccurate. Only 34% of data on materials could be used to calculate A1–A3 carbon emissions effectively. This gap between available information and valuable, high-quality data raises huge challenges for the construction industry. If these issues remain unaddressed, the ability to accurately meet safety, sustainability, and cost management goals will continue to be compromised.

Despite the significant data quality issues identified, the report outlines four steps that could solve 95% of the problems in waste and delivery documentation.

By adopting the outlined solutions and adhering to the seven principles of good data, construction professionals can ensure better outcomes. High-quality data will enable projects to stay on time, within budget, and meet regulatory requirements, ultimately creating safer and more sustainable buildings.

Improving data quality is not just a technical necessity; it is a fundamental requirement for the future success of the construction industry.

Table of contents

The State of Data Quality in Construction – Key Findings

WHAT DATA DID WE REVIEW?

Six years of data, from March 2018 - October 2024.

445 construction (new build and refurbishment) projects throughout England, Wales, Scotland, Australia, and the USA

434 UK projects

9 USA projects

2 AU projects

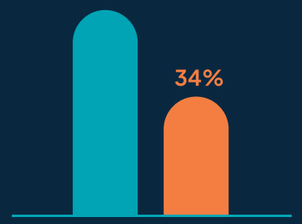
831,206 products across 363,608 delivery movements.

195,178 waste removals

Overall, it breaks down findings from **4,292,013 tonnes of products** and **4,169,067 tonnes of waste.**

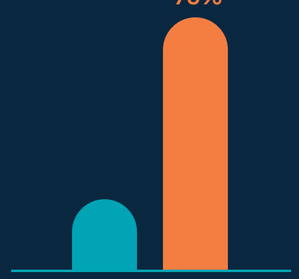
91%

of product and waste documentation needs to have data enriched in some way in order to make the data valuable for onward analysis.



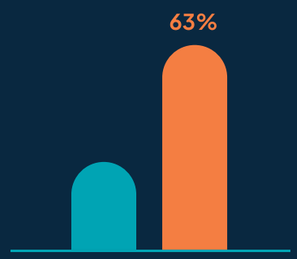
34%

of all construction materials could be used for effectively calculating product-stage carbon emissions (Modules A1–A3 of a whole life carbon assessment).



78%

of documented delivery movements tracked had enough data to calculate construction stage transportation carbon emissions (Module A4 of a whole life carbon assessment).



63%

of all waste removal documents could be used to contribute towards calculations of carbon emissions during all construction processes up to completion (Module A5 of a whole life carbon assessment).

Figure 1: The State of Data Quality in Construction - Key Findings

The State of Data Quality in Construction – Key Findings

COMPLETENESS

72%

of key construction materials were missing weight or volume information in their documentation

COMPLETENESS

95%

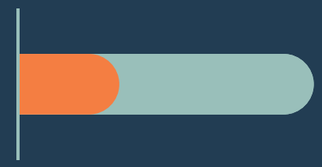
of total recorded deliveries were incomplete

UNIQUENESS

5%

of all delivery and waste data were duplicates

VALIDITY



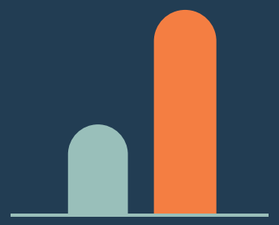
30%

of waste removal documents reviewed had invalid formatting of addresses

VALIDITY

14%

of product delivery documentation had invalid formatting of addresses



60%

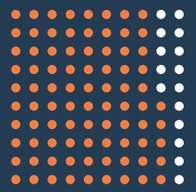
of waste removals did not have an associated weight or volume documented

ACCURACY

5%

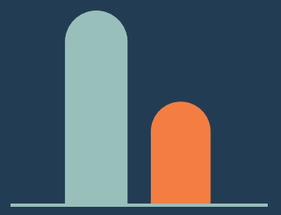
of waste data did not accurately reflect the true destination of waste removal

CONSISTENCY



85%

of data needed standardisation, to create data consistency.



32%

of waste removal documents reviewed had invalid Standard Industrial Classification (SIC) codes

AVAILABILITY

80%

of contractors don't have a structured way of tracking delivery data

Figure 2: The State of Data Quality in Construction - Key Findings

INTRODUCTION

Data quality is the vital foundation of informed decision-making in the Architecture, Engineering and Construction (AEC) sector. High-quality data is the only way to understand a project's true performance against sustainability and safety requirements, as well as internal cost and efficiency constraints that those in the sector have to manage.

With digital technologies rapidly being adopted across the AEC sector in the past two decades, the volume of data on any given construction project has skyrocketed. One study reported that the volume of available data in construction doubled between 2019– 2022 alone (Autodesk Construction Solutions). Despite this massive increase in data quantity due to digitisation, improving its quality often remains an afterthought.

This is particularly important due to the rising interest in implementing solutions powered by Artificial Intelligence to improve data automation, management and analysis. The concept "garbage in, garbage out" (GIGO) is very apt here. GIGO is used to highlight that poor-quality input data inevitably produces an outcome of similarly poor quality. Since Artificial Intelligence relies on the quality of the data inputted to produce effective results, an abundance of poor-quality data, especially within a limited data source, means incorrect and ill-informed decisions can be made.

This broad lack of consideration of data quality poses a significant risk to the industry.

AI algorithms rely heavily on the quality of data they process; any inaccuracies or delays can lead to outdated or incorrect insights.
Phanos Hadjikyriakou, CEO, 2050 Materials

Data is, of course, crucial to inform the design, build, and delivery stages of a project:

Architects and planners need high-quality data to gather key details about the material specifications so they can design a building that meets specific standards.

Project managers need high-quality data to judge whether they are on track to complete a project on time, budget and to specification.

For developers, high-quality data is vital for effectively managing a building post-completion and for retaining and enhancing its value for future sale.

Poor data quality has countless risks that can cascade into a host of larger problems with more serious consequences for the construction industry. Inadequate data can undermine the effectiveness of any data-based decisions made by professionals in the AEC sector at any level. A lack of knowledge of exactly what is happening on a construction site carries a host of significant safety, regulatory and financial risks. For example, high-quality data is vital for projects to stay within budget and schedule. Bad data in global construction may have cost $1.84 trillion due to poor decision-making in 2020 alone (Autodesk Construction Solutions).

Those projects that do have bad data do end up suffering in a disproportionate way.
Dev Amratia, CEO, NPlan

Due to the nature of the construction industry, low-quality data on delivered materials often has some of the most severe implications. When principles of high-quality data are not maintained, it can have catastrophic consequences. At worst, this can contribute towards disasters such as:

- The 2017 Grenfell Tower disaster, which claimed the lives of 72 people.
- The 1999 Viale Giotto 120 Building collapse in which 67 people were killed.
- The 2013 Rana Plaza collapse, which reached a death toll of 1,134 people.
- The 2021 Champlain Towers South condominium collapse, which caused the deaths of 98 people.
- The Val di Stava Dam collapse in 1985, which killed 268 people
- The Walkways collapsed in the Hyatt Regency Hotel in 1981, which caused 114 people to be killed.

In all of these cases, information about the true state of a building project not being shared transparently up and down the supply chain, particularly surrounding the materials used, the design specifications and safety compliance, led to horrifying outcomes. Low-quality data on materials left a void in valuable information, resulting in reduced accountability around who was responsible for ensuring safety was prioritised.

It is essential that those responsible for designing buildings have access to reliable information about the materials and products they wish to use.
The Grenfell Tower Inquiry (September 2024)

In addition, without high-quality data on delivered construction materials, it is impossible to understand compliance with key industry safety standards and sustainability assessment schemes such as BREEAM, LEED, and NABERS. Poor quality data on materials can also undermine the validity of whole life carbon assessments (WLCAs) that support reporting for regulatory and customer sustainability expectations.

Figure 3 demonstrates the various components of whole life carbon assessments, taken from the RICS Standards, and the breadth of data needed to calculate this. We refer to these life cycle stages throughout the report, highlighting A1–A3, A4 and A5 specifically. For further information on this, click here.

Data quality in construction is important because it ensures accountability. Accountability is what we need to transition the industry to a more climate-positive future
Phanos Hadjikyriakou, CEO, 2050 Materials

Maintaining and updating a digital audit trail of information relating to building work and maintenance throughout a building's life cycle is also a legal requirement in the UK under the Building Safety Act 2022. All stakeholders up and down the value chain must maintain a "Golden Thread" of information on these areas or risk not complying with legal duties around ensuring building safety.

This report provides an analysis of the state of data quality in the construction sector, specifically focusing on as-built materials and waste data and addressing the reporting, cost and regulatory challenges that frame the value of quality data in construction.

To achieve this, Part 1 outlines seven principles for data quality (Figure 2) and reviews 831,206 product deliveries to construction sites and 195,178 waste removals against these criteria.

Part 2 explores how to solve the data quality issues that arise. It outlines four steps that would have solved up to **95%** of the problems associated with the delivery and waste documentation we analysed.

In short, the report highlights the need to improve data quality in the sector and provides guidance to help construction sector professionals make this happen. It aims to support better data quality to enable buildings to become safer, projects to be delivered on time and budget, and environmental and regulatory risks to be reduced.

Figure 3: Whole Life Carbon Assessment for the Build Environment

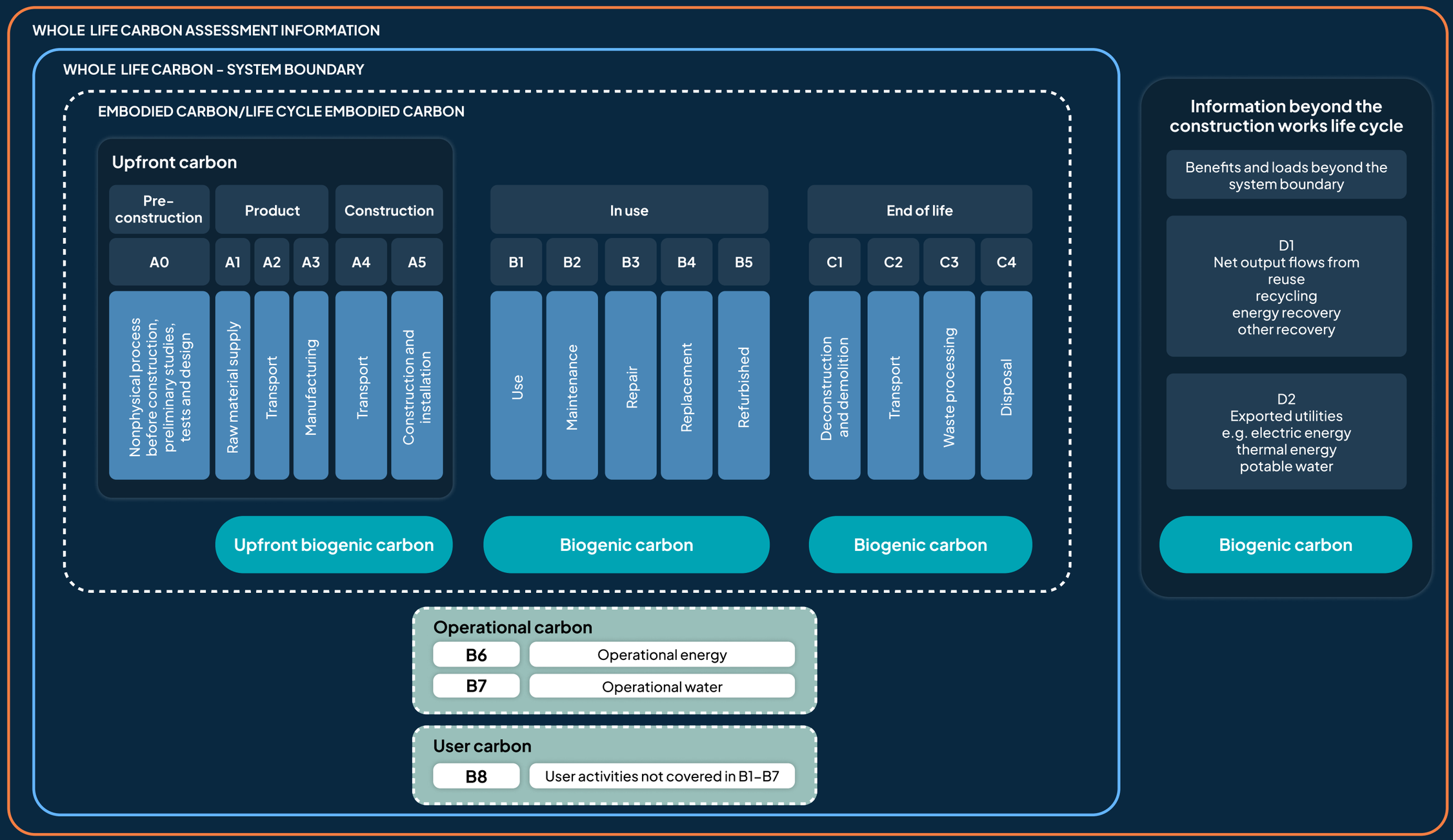


Figure 3: RICS Professional Standard, Whole life carbon assessment for the built environment, original graphic can be found at https://www.rics.org/content/dam/ricsglobal/documents/standards/Whole_life_carbon_assessment_PS_Sept23.pdf

THE FOUNDATIONS OF THIS ANALYSIS

What data was reviewed?

TO ASSESS THE STATE OF DATA IN THE CONSTRUCTION SECTOR FOR THIS REPORT, WE HAVE REVIEWED:

Six years of data, from March 2018 - October 2024.

445 construction (new build and refurbishment) projects throughout England, Wales, Scotland and the USA

- **434** UK projects
- **9** USA projects
- **2** Australian projects

831,206 products across 363,608 delivery movements.

195,178 waste removals

Overall, it breaks down findings from 4,292,013 tonnes of products and 4,169,067 tonnes of waste.

Due to the type of data used for this exercise, the description of data quality in the report focuses on material flow data. Although organisations may have greater data quality controls in place for some data sets, for example, safety or financial data, much of this still relies on manual reporting across a fragmented supply chain. Therefore, the risks and opportunities identified through the analysis of construction material data can be reasonably considered to be representative of the industry at large. As a result, the report assumes these extensive materials and waste data points are representative of the overall picture of data quality across the industry as a whole.

The specific data interpreted for this report includes the delivery documentation for materials entering construction sites and Waste Transfer Notes (WTNs), which track waste being removed from these projects. By utilising evidence of genuine material quantities as they move through the supply chain, these two sources of information enable a reliable reflection of as-built data, e.g. data that corresponds directly to buildings once they have been built.

Delivery documentation:
often contains a description or name for the products being delivered, information about the dispatch of these items, quantities, and suppliers.

Waste Transfer Notes (WTNs):
contain the classification of waste items in line with the European Waste Catalogue codes, quantities, locations of where the waste is taken from and moved to, and permit and licence details.

WHAT DOES ‘GOOD’ DATA QUALITY ACTUALLY MEAN?

Various frameworks and approaches exist to guide data quality evaluation depending on the specific needs of a particular organisation or industry. In compiling the following list of data quality principles, we have drawn from respected sources and selected the key dimensions we believe are most relevant to explore for the purposes of this report.

Principles of data quality

Accuracy
how well does the data reflect reality?

Completeness
is all the required data needed for something present and available?

Uniqueness
are there duplicates in the data?

Consistency
do data values conflict within another record or across different datasets?

Validity
to what extent does the data conform to expected formats, types, and ranges?

Availability
is the data accessible to those who need it and provided in a timely manner? Is the data up to date?

Valuable
is the resulting data informative and accessible to enable data-driven decision-making?

THE STATE OF DATA IN THE CONSTRUCTION SECTOR

What was the quality of the data reviewed?

ACCURACY

How well does the data reflect reality?

The accuracy of a data set can be determined by detecting anomalies that fall outside acceptable limits or do not match expected values. This is often the result of errors when collecting and recording data, including simply inputting the wrong number or selecting an incorrect category.

Findings:

Of the 140,738 waste items associated with a facility permit, 95% were accurate and fell within the expected bounds.

This means that 5% were inaccurate and did not reflect the true nature of activities on site.

In this example, we assessed accuracy by reviewing the legitimacy of waste facility permit IDs and licences on waste documentation, of which 5% were identified as inaccurately logged on waste transfer notes.

What could this mean?

Having data that does not reflect the realities of products being delivered or waste being removed means that uninformed decisions are being made and inaccurate reports are being submitted. This could impact project management, safety and compliance.

Without knowing exactly what quantities and types of materials were used in a building, it is not possible to assure the safety and appropriateness of materials, or calculate its whole life carbon footprint reliably.

Accurately assessing costs and forecasts for future bids will be undermined if the information used is inaccurate.

In the above example, the inaccuracy of waste facility permits and licences logged on documentation could have resulted in over £2.1 million of potential fines from regulators such as the Environment Agency.

You can collect as much data as you want, but if it's not accurate, it's pointless.
Tony Harbour, Director or Partnerships EMEA, Procore Technologies

COMPLETENESS

Is all the required data needed, present and available?

Measuring the completeness of a data set involves the identification of records with empty or incomplete fields. Examples of incomplete data we found when assessing product delivery documentation included documents missing product names, descriptions, metrics like weight or volume, and details such as dispatch addresses. Incomplete data is not always entirely missing but can refer to the lack of any meaningful information.

Findings:

Across the 831,206 product delivery documentation reviewed, 95% of total recorded deliveries were incomplete.

The completeness percentage of this data was, therefore, only 5%.

In fact:

- 6% of product deliveries had non-descriptive product names or descriptions. This was determined by products of which over 50% of the name or description contained symbols or numbers, e.g. UC15215223.
- 72% of key construction materials delivered were missing weight or volume information in their documentation.
- 18% (almost 1 in 5) of all delivered goods were missing a dispatch address.

Also, in reviewing 195,178 waste removals:

- 60% did not have an associated weight or volume documented.

Key construction materials refer to the products that typically make up the majority of materials used for constructing a building or infrastructure project and those which are often the most carbon-intensive: Insulation, Timber, Structural steel, Plasterboard, Bricks & Blocks, Concrete, Aggregate, Fuel and Rebar.

What could this mean?

The non-compliance risk caused by having 60% of waste removal data without weight or volume documented is significant. The 2023 UK Construction Industry Waste report found over £23 million of possible waste data-related fines unclaimed by the Environment Agency on just 90,000 waste transfer notes.

More broadly, key details being absent make the resulting data set difficult to interpret and use for onward reporting and analysis. A lack of complete data can also impact the day-to-day project management on a construction site, impacting cost and efficiency key performance indicators.

For example, assessing if the correct material is arriving on-site is not possible with product descriptions and names that are difficult to interpret.

Calculating as-built A1–A3 embodied carbon cannot be completed if key information on specific materials, such as volume and/or weight data and types of materials delivered, is absent.

A lack of complete data can also impact the day-to-day project management on a construction site, impacting cost and efficiency key performance indicators.

The barriers to perfection, actually, all stem around effort and cost.
Greg Lawton, CEO, Nodes & Links

UNIQUENESS

Are there duplicates in the data?

Data uniqueness tracks duplicate data points to see if any pieces of information are being counted more than once. Duplicate data entries may even have certain fields with unique information when mistakes are present in the duplicate, potentially compounding the issue.

Findings:

Across over a million pieces of data reviewed for this report, 5% were found to be duplicates.

What could this mean?

Having duplicates in your data set can unduly weight the results, making values seem higher than they are in reality. Insights and analysis will, therefore, be based on incorrect information.

Duplicate data entries could indicate false project progress if it's assumed that more material has been brought onto site than has actually happened in reality.

As embodied carbon targets come into force, projects could be unfairly penalised for higher than anticipated embodied carbon due to double counting of material consumption and/or waste production. In February 2023, EU carbon prices reached a record high of €100.34 per metric ton of CO_2 (Statista, 2024). Therefore, if carbon offsetting or insetting is being utilised, the impact on a building could be thousands of euros if the total embodied carbon figure is off by only 1%

CONSISTENCY

Do data values conflict within a record or across different datasets?

Data consistency involves assessing whether any data entries conflict with each other, within or between different datasets. This is a particular issue when combining multiple data sources.

Findings:

Across the data reviewed, 12,814 different company names were listed as suppliers of products to construction sites, but, in fact, only 6,917 of these were unique companies.

Therefore, inconsistencies in naming conventions across documentation resulted in an 85% increase in additional unnecessary information, which was not actually unique.

What could this mean?

When data is structured and input in an inconsistent manner, conflicts are created, which undermine efforts to interpret and use the data for decision-making. Inconsistencies between duplicate data points can also make it difficult to spot the repetition, compounding any data uniqueness issues.

Identifying risks such as underperforming suppliers will not be effective if a supplier is represented by multiple different names.

By having better quality data, you're going to be able to drive better outcomes on the construction site because it enables you to make better-informed decisions on the task you're trying to achieve.
Tony Harbour, Director or Partnerships EMEA, Procore Technologies

VALIDITY

To what extent does the data conform to expected formats, types, and ranges?

Data validity measures the extent to which data conforms to expected formats, types, and ranges. Data entered into a specific field needs to exist within the appropriate boundaries; for example, months of the year entered as numbers should never exceed twelve. If the data is not in the right format, it is invalid and may be unusable.

It is important to note that data validity and data accuracy, while similar, should not be confused for being the same data quality issue. A data entry may be within valid bounds but may not be accurate.

Findings:

Of the 195,178 waste removal documents reviewed:
- 30% had invalid formatting of addresses
- 32% had invalid Standard Industrial Classification (SIC) codes
- 1.62% had invalid European Waste Catalogue (EWC) codes

Across the 831,206 pieces of product delivery documentation:
- 14% had invalid formatting of addresses

What could this mean?

Data that is invalid is often simply unusable if it does not fit into predetermined or expected boundaries that act as the foundation for understanding. Invalid data leaves a void of information that would otherwise be useful and can be hard to spot without knowing the usual bounds.

Invalid EWC codes will reduce how effectively regulatory risk can be managed, and means that waste recovery routes cannot be accurately calculated.

If there is a link broken within data collection, it will create anomalies within the outputs and the outcomes that you want to drive.
Tony Harbour, Director or Partnerships EMEA, Procore Technologies

AVAILABILITY

Is the data accessible to those who need it, and provided in a timely manner?

Data availability refers to how accessible new data is to those who need it, particularly focusing on its timeliness. A lack of timely and up-to-date information on any given topic would mean there is low data availability in this area.

Findings:

Of the 445 projects analysed, most Tier 1 contractors collecting waste data via manual or traditional processes had a previous time lag of a month or more to obtain data from their carriers and brokers.

Approximately 80% of these contractors didn't have a structured way of tracking delivery data, instead relying on expected and assumed amounts or basing their analysis on spend data, which is generally considered suboptimal for carbon reporting.

What could this mean?

The availability of data is vital to ensure that information can actually be used to inform decisions. If the data is not provided in a timely manner, it risks being out of date. This can cause decisions to be made based on out-of-date information, which incurs a host of safety, regulatory and cost risks.

A delay in receiving accurate information about the materials delivered to a site can result in the installation of inappropriate materials requiring costly re-work at a later date or potentially introducing serious risk to inhabitants down the line.

In construction, data is often not updated regularly enough to keep up with the pace of change.
Phanos Hadjikyriakou, CEO, 2050 Materials

VALUABLE

Is the resulting data informative and accessible to enable data-driven decision-making?

Data value refers to how useful data is for enabling data-driven decision-making. If data is not collected and used with defined purposes in mind, it is at risk of being redundant. The valuable nature of data should, therefore, be considered when assessing the overall quality.

Data is the cornerstone of decision making.
Dev Amratia, CEO, NPlan

Data itself is not a valuable resource, it's what the data enables you to do that puts value to the data.
Greg Lawton, CEO, Nodes & Links

Findings:

In order to make the data reviewed usable and valuable for onward analysis, 75% of all waste documentation and over 95% of all documented material deliveries needed to have data enriched in some way.

Onward analysis relates to overall waste reporting, carbon accounting, material flow analysis, construction management and programme risk analysis.

For example, with respect to carrying out whole life carbon assessments (WLCA):

- **Only 34%** of all construction materials could be used for effectively calculating product-stage carbon emissions (Modules A1–A3 of a whole life carbon assessment).
- **Only 78%** of documented delivery movements tracked had enough data to calculate construction stage transportation carbon emissions (Module A4 of a whole life carbon assessment).
- **Only 63%** of all waste removal documents could be used to contribute towards calculations of carbon emissions during all construction processes up to completion (Module A5 of a whole life carbon assessment).

See Figure 3 for a breakdown of WLCA.

What could this mean?

The absence of high-quality data reduces how usable and, therefore, valuable it can be for decision-makers within the AEC sector. Data value is the most important metric of overall data quality; if the information collected is not valuable for its intended purpose, or in fact, if it serves no purpose, it prevents effective action from being taken.

It is true that 'you can not manage what you do not measure', but it is also important to remember that **the point in measuring is to be able to manage**, not just report.

Achieving Net Zero Carbon status for new buildings will be difficult without due consideration for data quality and verification.

Passing Gateway 3 of the Building Safety Act will be difficult without the evidence for as-built materials, part of the Golden Thread of data.

Managing costs, safety and sustainability considerations in real-time will be impossible if data does not include information that makes it valuable for these ends.

The majority of the construction industry has data on their projects that is difficult to use.
Phanos Hadjikyriakou, CEO, 2050 Materials

As you want more and more capability, the quality of your data has to increase, probably exponentially.
Greg Lawton, CEO, Nodes & Links

A Whole Life Carbon Assessment (WLCA) is a comprehensive framework for evaluating the carbon emissions of buildings and infrastructure throughout their entire lifecycle. WLCAs follow a modular structure for carbon reporting, which breaks down the built asset's life cycle into stages and modules.

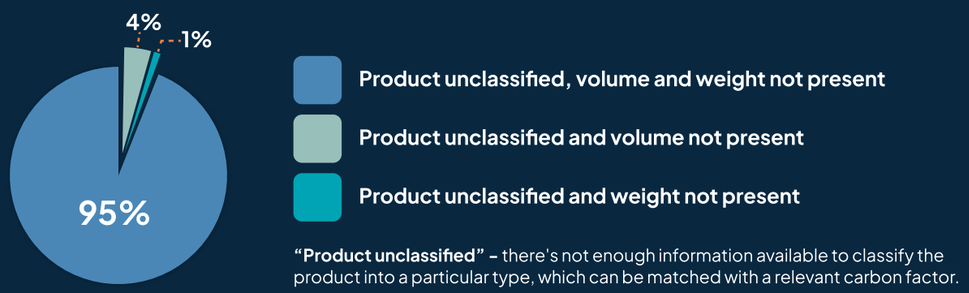
According to the WLCA Standard from the Royal Institution of Chartered Surveyors (RICS), modules A1–A3 refer to product-related carbon emissions and modules A4 - A5 refer to carbon emissions from the construction stage.

What prevents the data from being valuable for calculating modules A1 – A3?

The most common reason for materials being excluded is the **missing weight/volume data**, which is required in order to calculate the overall embodied carbon, even if an appropriate carbon factor can be matched.

Products lacking names or meaningful descriptions create challenges around classifying products. This limits the way in which they can be assigned the correct carbon factor.

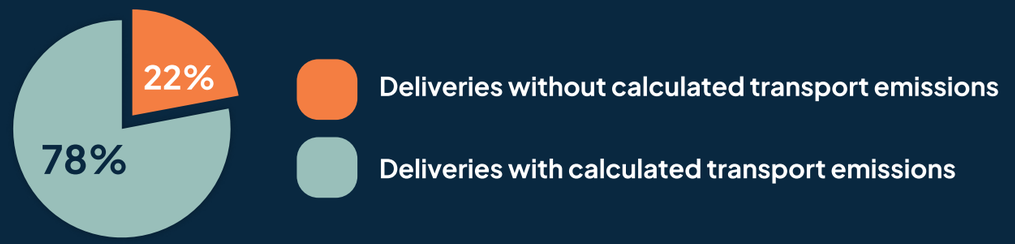
What's preventing us from populating the 66%



"Product unclassified" – there's not enough information available to classify the product into a particular type, which can be matched with a relevant carbon factor.

What prevents the data from being valuable for calculating module A4?

In terms of calculating module A4 of a whole life carbon assessment, an absence of either the dispatch or delivery addresses on the documentation limits the value of this data. This is because addresses are required to help calculate the overall transport emissions associated with deliveries.



What's preventing us from populating the 22%

What prevents the data from being valuable for calculating module A5?

Waste data lacking value for helping to calculate module A5 of a whole life carbon assessment is the result of a host of factors, including:

71% of waste collections were missing weight and volume information

13% of waste collections missing EWC codes

34% of waste collections missing facility permit

Figure 4: What prevents the data from being valuable for calculating modules A1 - A5

WHAT RISKS DOES POOR QUALITY DATA POSE?

Poor data across the seven principles outlined in this report will permeate every construction project stage, creating a range of risks.

Compliance Risk

Inaccurate or missing data can lead to non-compliance with building codes, environmental regulations, or safety standards, resulting in legal penalties or financial forfeiture.

The “Golden Thread” is a legal requirement under the Building Safety Act 2022. That acts as a digital audit trail, which is updated throughout a building’s life cycle.

All stakeholders up and down the value chain must now collect and share data that is reliable, consistent and accurate or be at risk of not complying with legal duties and not ensuring building safety.

Higher-risk buildings, such as high-rise apartment buildings, must be designed and built with utmost care and attention to detail. The “Golden Thread” ensures that decisions made throughout the life cycle of a building use information that can be depended upon. High data quality ensures that the foundations of these critical decisions are robust, reducing errors and improving safety.

Safety Considerations

Incorrect data on site conditions, materials, or structural elements can increase the likelihood of accidents or safety violations, endangering the lives of workers and future inhabitants.

The Grenfell Tower disaster is an example of data and information mismanagement undermining safety considerations with fatal consequences.

The disaster was the deadliest residential fire since WWII, claiming 72 lives in June 2017. A public enquiry later showed that the construction company had used cheaper, flammable cladding on the building. It highlighted failures in construction product regulation and emphasised that every party responsible for ensuring the safety of the building failed in some way, and all deaths were ultimately avoidable. This event was the catalyst for the new Building Safety Act 2022 and Fire Safety Act 2021.

Risk of Financial Loss

Low-quality data can misallocate resources and inaccurate budgeting, resulting in cost overruns and financial losses.

Organisations spend between 10–30% of revenue on handling data quality issues (DAMA Data Management Body of Knowledge).

Ineffective Sustainability Assessment

Missing data can hinder the thorough evaluation of environmental impacts and sustainability performance, making it challenging to meet sustainability standards such as BREEAM, LEED or NABERS.

For example: looking at Figure 4, the missing 66% of the picture when calculating product-stage carbon emissions and not having access to 22% of delivery movements or valuable waste data for construction-stage emissions means that companies could be significantly under-calculating their as-built upfront carbon emissions.

WHAT IS BREEAM?

The Building Research Establishment Environmental Assessment Method is a UK-developed sustainability assessment method for buildings, infrastructure, and masterplan projects. BREEAM evaluates the environmental performance of a building across various categories, such as energy use, water consumption, indoor environmental quality, and material usage.

WHAT IS LEED?

Leadership in Energy and Environmental Design is a green building rating system developed in the United States. The latest LEED certification, LEED v5, will launch in the first quarter of 2025 and place a strong focus on lower embodied carbon approaches.

WHAT IS NABERS?

The National Australian Built Environment Rating System is a sustainability rating system that assesses the environmental impact and energy efficiency of buildings. It provides a rating of one to six stars based on a building's performance across: Energy, Water, Waste, Indoor environment, and Carbon emissions.

Failing to meet industry emissions targets

Data is a fundamental part of tracking carbon budgets and performing the whole life carbon assessments (WLCAs) that are vital to track progress towards industry net zero goals.

A whole life carbon assessment (WLCA) is a method for measuring the carbon emissions of a building produced throughout its entire life cycle, from the construction, use, and end-of-life stages, including both embodied and operational carbon.

WLCAs enable developers to meet government and client demands around reducing embodied carbon and creating net-zero buildings. Without data being collected at every stage of a building's life cycle, there is no possible way to measure and manage carbon emissions in a reliable and consistent manner that is needed for completing a WLCA.

Failure to undertake an accurate WLCA will potentially result in reputational damage and a restriction in a company's client base as more developers shift their business to those with verifiable carbon assessments.

Impacting the quality of a build

Bad data related to design or material specifications can compromise the quality of the build, leading to defects, rework, or failed inspections.

14% of all rework in construction globally is caused by bad data (Autodesk, 2020).

Project Management Risk

Procurement and delivery data errors can result in materials not arriving on time or the wrong materials being delivered, disrupting project flow, creating time lags and increasing costs.

> "Incorrect data with no data integrity checks creates a domino effect of cascading errors with wrong analysis, leading to wasted time." — Brooke Wandler, Head of Growth Marketing, Nodes & Links

Time lags

Time spent correcting bad data or verifying information can significantly delay project progress, impacting schedules and pushing back project completion.

> "Contractors can spend up to 1 day per week manually validating their supply chain data to ensure it is high quality." — Tier 1 Contractor

> "Previously all our waste data had to be manually entered into an internal reporting system. The site team would manually upload every single waste transfer note from the demolition, excavation, and construction phases of the project. Ensuring each waste transfer note was compliant with the Environment Agency's waste duty of care regulations was also a time-consuming process." — Tier 1 Contractor

Bridging the industry gap between available information and valuable data is vital to improve all aspects of the AEC sector that rely on information for:

- Setting strategy and direction during the planning and design phases.
- Managing the project's success during the construction phase, including ensuring compliance with necessary standards.
- Assessing the whole project against key performance indicators after project completion.

WHAT CAN THE INDUSTRY DO TO IMPROVE ITS MATERIALS AND WASTE DATA QUALITY ACROSS THE SUPPLY CHAIN?

It is clear from the findings in Part 1 that while there may be abundant construction-related data, the quality of materials and waste data is generally poor. As discussed above, these quality issues can increase various risks and reduce the value of the data for meeting crucial responsibilities, from decision-making or carbon reporting.

Improving the quality of data is, therefore, a productive step in improving the quality of decision-making and reducing any risks that may otherwise occur.

Four steps for improving the quality of materials and waste data

Improving data quality doesn't require a seismic shift in the way we collect, store and manage data. Implementing a few simple changes can have a great impact.

Delivery documentation

1 Include descriptive product names that can be easily interpreted by a human and allow for correct classification.

Bad	Good
LX5-7W-3000K-D	2440 x 1220 x 15mm Elliotis Pine Plywood C+/C CE2+ EN636-2S EN314 Class 3 E1 FSC® certified - (60)

2 Include weight and/or volume for all products and materials being delivered.

3 Include the address details for each delivery (dispatch and delivery addresses).

Waste documentation

4 Include the key data requirements*, such as:

- A description of the waste
- The container for the waste
- Quantity of waste
- Date of waste movement
- Site address of waste collection and carrier address
- Carrier licence
- Facility permit or waste exemption relating to waste destination**
- European Waste Catalogue (EWC) code and the Standard Industry Code (SIC)

*Note: as per the Waste (England and Wales) Regulations 2011.
**Note: this is the most common missing information, and critical to include in order to understand the waste recovery routes of each waste movement, alongside being a legal requirement.

Figure 5: Four steps for improving the quality of materials and waste data

If construction professionals followed these four steps relating to supply chain data, up to **95% of the problems** found in relation to data completeness for within the waste and delivery documentation could **be solved**.

For even better data quality...

- **For improved classification** according to industry standard systems, such as Uniclass, OmniClass, MasterFormat and others, it is crucial to include detailed descriptions within product names. Ideally, cataloguing these products will enable traceability. This practice not only aligns with industry standards but also reduces the potential risk associated with misclassification errors further down the supply chain.

- **For composite products,** it is important to include descriptions that break down details of each individual component or raw material. A way to achieve this data enrichment would be by providing details that link back to other documents in open supplier databases such as Environmental Product Declarations (EPDs). Doing this will give some insight into the constituent materials, enabling the verification of the products arriving on-site compared to those in the design. This is vital for overall quality management and will be useful for carbon reporting estimations if the original data cannot be matched with relevant carbon factors.

HOW CAN DIGITAL SOLUTIONS HELP IMPROVE CONSTRUCTION DATA QUALITY SURROUNDING MATERIALS AND WASTE?

Digital technologies are transforming data collection and management across the AEC sector by enabling improvements across the seven data quality principles reviewed in Part 1 (accuracy, completeness, uniqueness, consistency, validity, availability and value). These advancements can help reduce errors, enhance transparency, and streamline decision-making processes.

Qflow's solution is one example of an innovative technology improving the quality of materials and waste data. The following section explores how this can improve data quality across the seven principles.

ACCURACY

How well does the data reflect reality?

Qflow improves accuracy by correcting false or out-of-date data from documentation. It does this by referencing data from external sources to ensure the legitimacy of documentation, such as data from the Environment Agency to cross-reference waste facility permit IDs and licences.

RESULTS:
Across the full data set, Qflow auto-corrected the 5% of inaccurate waste information by using data from reputable external sources to validate information logged on waste documentation.

COMPLETENESS

Is all required data for a particular use present and available?

Qflow can improve data completeness by gap-filling data that isn't fully included within the documentation. It does this by referencing estimates from its global dataset and finding points of similarity to ensure key pieces of missing information are provided.

RESULTS:
Since its inception, Qflow's technology has gap-filled weight data for 20% of recorded deliveries to construction sites and 62% for waste removals.
Overall, this represents a 41% increase in total recorded weight, which would have otherwise been missing.

Qflow's system can also gap-fill 75% of waste facility information, using the Environment Agency as a key source. This improves data completeness for important data required to meet compliance on waste documents.

BERKELEY

Qflow's platform helped Berkeley identify missing data required for legal compliance and for internal reporting, such as missing facility permit number, container types, or materials certification. This automated process reduces the time required to manually check the documentation, and allows us to work with our supply chain to address any non-compliance at the source, thus preventing that from re-occurring in future. This minimises the risk of breaches in regulations, ensures we have the required information for our reporting and supports our broader push towards environmental leadership.

WILMOTT DIXON

For key construction materials, Wilmott Dixon identified and 'completed' the following data in order to use it for further reporting such as doing a Whole Life Carbon assessment.

- **They found that 63% of key construction material tickets were missing information on weight.**
- With Qflow, they reduced this number by **45%**.

UNIQUENESS

Are there duplicates in the data?

Qflow automatically checks for duplicate records across datasets and highlights them during internal quality control processes.

RESULTS:
At the time of writing, 1% of Qflow's customers' data was identified as duplicates. By identifying duplicates and auto-voiding these where possible, Qflow's technology removes unnecessary data 'noise', allowing for easy decision-making.

CONSISTENCY

Do data values conflict within a record or across different datasets?

Qflow regularly checks and standardises company names, including supplier and contractor names.

RESULTS:
Qflow has reduced the total name variation across all data it has handled to date by 46%. This limits the redundant variety within data and allows for a more standardised and reliable dataset.

VALIDITY

To what extent does the data conform to expected formats, types, and ranges?

There are several methods by which Qflow can provide useful estimates and gap-filling for invalid data. For example, if product deliveries are recorded with symbols or numbers rather than names or letters, Qflow can use wider information from the delivery to estimate the contents.

RESULTS:
Despite the low level of definition with some product deliveries reviewed in this report, for example, where the product names contained more than 50% symbols or numbers, Qflow was still able to categorise 68%.

In addition, 2% of the waste transfer notes had invalid European Waste Catalogue (EWC) codes. With Qflow, this was improved to 100% validity.

AVAILABILITY

Is the data accessible to those who need it and provided in a timely manner?

Qflow's live data capture technology speeds up data collection compared with manual processes.

RESULTS:
The time delay from deliveries/waste removals taking place on-site to having available data on these can be reduced from around 1 month to just 16 hours using Qflow.

This was calculated by taking the median live ticket publication time of Qflow users over a period of 1 month.

For around 80% of Qflow's customers, the platform also provided 'actuals' for their materials data for the first time, moving away from relying on estimated or spend-derived quantities.

VALUABLE

Is the resulting data informative and accessible to enable data-driven decision-making?

By improving various aspects of data quality, Qflow is able to boost how valuable data can be for certain tasks.

By gap-filling over 3 million tonnes of material and waste data using their global dataset, the embodied carbon calculations of Qflow customers are more complete. This is evidenced by a 41% increase in the amount of recorded weight on materials and waste documents.

This means that 97% of the key construction materials needed to quantify product-related carbon emissions (modules A1–A3 of a whole life carbon assessment) can be included in embodied carbon calculations and reporting where previously it could not.

In this case, the gap-filling process enabled by Qflow's digital technology improves how valuable datasets can be for such carbon reporting tasks.

We noticed a significant proportion of our Scope 3 carbon emissions coming from one supplier due to their location. Qflow helped us to identify 770 kgCO_2e that was wasted on a single delivery using that particular supplier.
Anonymous fit-out contractor

Using Qflow's platform highlighted incomplete and inaccurate data including missing facility permits, container types and addresses. This saved us over two thousand pounds in non compliance fines
Anonymous construction company

IMPROVING DATA VALUE

The nature of how valuable data is depends on the specific intended or desired purpose. Therefore, improving the value of data can take on many different forms for many different reasons. Below are two examples of outcomes resulting from Tier 1 Suppliers improving their data value.

IMPROVING SUSTAINABILITY COMPLIANCE

Galliford Try, a Tier 1 Contractor, implemented Qflow on its £85 million Melton Mowbray Distributor Road project, commissioned by Leicestershire County Council. This partnership streamlines the collection of data across more than 10 direct suppliers and their extended supply chains. Since the project's inception, Qflow has captured over 10,000 product deliveries and waste transfer notes, facilitating more efficient monthly carbon data reporting to National Highways. This unified data approach has minimized manual processing, enabling Galliford Try to compile and report carbon data more swiftly and accurately.

BW Construction, a Tier 1 Supplier, required a low carbon design and build for their re-design of an office space across 4 floors.

Using Qflow, the site team were able to log every material delivered by taking a photograph of the delivery note. This improved the completeness of their dataset and increased its value for tracking Material and Waste BREEAM credits in real-time.

Results of using the platform included:
- 26 risks highlighted
- 100% compliant FSC timber
- Reducing waste intensity to less than 3.5 tonnes/100m² GIA (gross internal area). This is well below the industry best practice recognised by BREEAM of 6.5 tonnes/100 m² GIA.

WILL IMPLEMENTING DIGITAL TECHNOLOGIES SOLVE ALL DATA ISSUES?

Moving from paper-based systems to digital ones for collecting and managing construction data does not always guarantee an improvement in data quality. This is because digital technologies cannot always rectify the underlying issues resulting from data input.

Factors holding back data quality include:

- **Lack of standardisation and interoperability:** Construction companies use a variety of different platforms, tools, and approaches for data, which leads to challenges in integrating data and information across systems.

- **Data security issues:** Many stakeholders view their data as proprietary and are reluctant to share it up and down their supply chain without assurances of security and ownership.

- **Lack of data literacy:** The construction industry faces a shortage of professionals skilled in collecting, managing and analysing data. This increases the likelihood of data being mismanaged and underutilised, especially when manual input is required in digital systems.

- **Industry silos:** The fragmented nature of the construction industry means that those at the source of data collection are not motivated to capture data that is only relevant to others further up the supply chain.

Methods must, therefore, be implemented to overcome data quality issues regardless of a company's given level of digitalisation.

These include:

1. **Data Standardisation:** Establishing clear, uniform standards for data collection across the construction sector and within project teams can reduce data-related inefficiencies.

2. **Rigorous Data Governance:** Introducing a framework on data governance can improve accountability and address challenges around intellectual property rights by outlining clear data responsibilities.

3. **Training and Skill Development:** Investing in training programs to educate industry professionals on the importance of data collection can increase data quality.

4. **Improving Collaboration:** Encouraging collaboration among different departments and stakeholders can help break down silos and motivate those throughout the supply chain to maintain data standards.

Thankfully, there are already a host of initiatives that focus specifically on these issues and aim to address them, such as:

- The Construction Data Trust is a not-for-profit that collaborates with the industry to securely share and manage construction project data. Its aim is to generate valuable insights from project data to reform how projects are managed and upskill the broader ecosystem.

- Uniclass is a unified classification system that standardises the organisation and structuring of construction information, improving consistency and communication across projects and stakeholders.

- The openBIM® approach, defined by buildingSMART International (bSI), aims to improve the accessibility, usability, management and sustainability of digital data. It directly combats industry inefficiencies by fostering better teamwork and new ways of working.

- The Construction Innovation Hub is a UK government-funded initiative developing user-friendly guidance, training and tools to encourage more organisations to benefit from data-driven decision-making and secure, resilient data sharing.

- The Information Management Framework (IMF) from the Centre for Digital Built Britain is a set of open, technical and non-technical standards, guidance, and resources designed to enable seamless, secure data sharing. It ensures resilient data interoperability while addressing security, legal, commercial, and privacy concerns.

CONCLUSION

This report highlights the state of data quality in the Architecture, Engineering, and Construction (AEC) sector, the implications of this and opportunities for improvement. Through the analysis of over **1 million product and waste documents**, the findings reveal significant issues with poor-quality data during the construction stage. Of the data reviewed, **95% of delivery documentation and 75% of waste transfer notes (WTNs) were found to be incomplete, inconsistent, or inaccurate.** This directly impacts the sector's ability to calculate and report on carbon emissions, meet sustainability goals, and adhere to regulatory requirements.

To address these issues, there are **four steps that can resolve up to 95% of the data quality problems identified**. These solutions revolve around making a few simple changes to data collection, namely, including key descriptions. By implementing these strategies, construction professionals can drastically improve data quality, enabling more precise decision-making, reducing costs, ensuring compliance and delivering on sustainability targets.

The issues identified with waste and materials data are not isolated but represent a broader problem with data management in the construction industry. If data quality at such fundamental stages of construction is not being treated seriously, then it is fair to assume that these issues permeate throughout different data types within the sector. Poor data quality for construction materials alone can undermine the sector's ability to manage safety standards, regulatory compliance and profitability. With such significant implications for just these two sets of data, **it is clear that the sector's approach to data quality needs improving.** The industry must address these overarching data issues or risk compromising future projects' sustainability, safety, and financial viability.

Improving data quality is essential for the future of the construction industry. The sector is currently in the process of a multifaceted transformation with digitisation accelerating, transparency increasing in importance and sustainability rising up the regulatory agenda. High-quality data will, therefore, become even more critical for decision-making and meeting regulatory demands. **Maintaining a thread of high-quality data throughout a building's lifecycle is becoming a vital requirement for project success** in terms of internal profitability, legal compliance, and adherence to sustainability goals.

Going forward, focusing on improving data quality will enable the construction industry to unlock the full potential of digitalisation. By addressing the underlying causes of poor-quality data, the sector can leverage digital technologies and innovative tools to ensure more efficient, safer, and sustainable projects. **The success of future construction projects depends not just on new technologies but on the accuracy and reliability of the data driving them.** This report hopes to have provided a pathway for making those improvements, laying the foundation for a more efficient, accountable and sustainable industry.

GLOSSARY

Architecture, Engineering, and Construction (AEC): A sector encompassing the disciplines to plan, create, and maintain the built environment, including buildings and infrastructure.

Carbon offsetting: Compensating carbon emissions by funding projects that reduce or absorb an equivalent amount of CO_2.

Construction stage carbon emissions (modules A4 – A5 of a whole life carbon assessment): Cover emissions related to transportation of construction products and all construction and installation processes, including wastage, up to project completion.

Composite products: Goods that are made by combining multiple elements (e.g. wood, metal, polymers) to create a whole new product for improved strength or functionality in construction applications.

Data interoperability: The ability of different software systems to exchange and use data seamlessly within projects, ensuring consistent and accurate information sharing.

Delivery documentation: Contains descriptions or names for the products being delivered, information about the dispatch of these items, quantities, and suppliers.

Embodied carbon: Carbon emissions associated with materials and construction processes throughout the whole lifecycle of a building or infrastructure.

European Waste Catalogue (EWC) codes: Standardised codes classifying waste types across the EU, enabling consistent tracking, handling, and reporting of waste.

Gap-filling: Process of addressing missing or incomplete information within a dataset, typically in project or compliance documentation.

Key construction materials: Products that typically make up the majority of materials used for constructing a building or infrastructure project and are often the most carbon-intensive: Insulation, Timber, Structural steel, Plasterboard, Bricks & Blocks, Concrete, Aggregate, Fuel and Rebar.

Net Zero: A balance where carbon emissions are reduced to the lowest levels possible, with remaining emissions offset to achieve zero overall impact.

Product-related carbon emissions (modules A1–A3 of a whole life carbon assessment): Emissions from raw material extraction, transport, and manufacturing stages of building materials in a lifecycle assessment.

Standard Industrial Classification (SIC) codes: Codes categorising industries and business activities for statistical, regulatory, and compliance purposes.

The Golden Thread of information: A digital record of a building's lifecycle information ensuring safety, accountability, and transparency for future reference and compliance, which forms part of The Building Safety Act.

The Building Safety Act: UK legislation aimed at improving building safety through stricter regulations, especially in the wake of the Grenfell Tower tragedy.

Tier 1 Contractors: Large, well-established construction companies responsible for managing and delivering large-scale construction projects, overseeing all aspects of project delivery while managing subcontractors.

Waste Transfer Notes (WTNs): Contain the classification of waste items in line with the European Waste Catalogue codes, quantities, locations of where the waste is taken from and moved to and permit and licence details.

Whole life carbon assessment (WLCA): A Method for measuring the carbon emissions of a building produced throughout its entire life cycle, from the construction, use, and end-of-life stages, including both embodied and operational carbon.

About Qflow

Qflow was founded with a vision to create a construction industry that uses only the resources it needs in the most efficient way possible – delivering a built environment that meets the needs of society without compromising future generations.

To help deliver this, Qflow's technology streamlines data capture on the materials used within buildings, future-proofing the industry's ability to recover assets at their end of life and providing information on the embodied carbon and risks of these materials today.

The platform enables contractors to collect data on materials and waste in close to real-time by extracting data from documentation on material movements on-site. This provides the basis for understanding the embodied carbon of new or refurbished spaces, as well as highlighting risks in the supply chain.

Given the importance of data to real estate delivery and management, this report is intended to shine a light on the current state of data quality within construction, using materials and waste data as an example by drawing on over 6 years of data collected through Qflow's platform to date.

Read more about our work

Qflow

THE STATE OF DATA QUALITY IN CONSTRUCTION

January 2025

Author
Qualis Flow Limited

Design
Adopter Limited

Acknowledgements
2050 Materials, NPlan, Procore Technologies, and Nodes & Links