

# The importance of data quality in the journey toward decarbonization

JUNE 2023

A co-authored report from:



**PCAF**

Partnership for  
Carbon Accounting  
Financials



**CDP**<sup>™</sup>

**DISCLOSURE INSIGHT ACTION**

# Table of Contents

- 1. **Background**..... 3
- 2. **Data quality: Introduction**..... 4
- 3. **CDP and PCAF data quality scoring**..... 5
  - What are the CDP Data Quality Scores? ..... 5
  - How were CDP’s Data Quality Scores developed? ..... 5
  - PCAF data quality scoring system ..... 5
- 4. **PCAF and CDP alignment** ..... 6
- 5. **Conclusion** ..... 6

# Background

One of the greatest challenges facing financial institutions when it comes to measuring emissions is the scarcity of high-quality data and to address this, the standardization of data quality scoring is essential.

For more than two decades, CDP has been working with investors, companies, and governments on an annual basis to drive industrial-scale environmental disclosure on climate change, water security, and deforestation.

These disclosures are then scored based on both the level of transparency and the strength of environmental actions that are disclosed. Importantly, all scores and all disclosure data are also made immediately available to the 750+ strong financial institutions that support CDP as “Signatories”. CDP harnesses these disclosures to create innovative data products such as the *Full GHG Emissions Dataset* and the *CDP Temperature Ratings*. These products, respectively, provide cleaned and modelled scope 1 - scope 3 emissions data for over 7,500 companies and display Implied Temperature Ratings with comprehensive targets data on 12,000 companies.

In 2021, CDP started its partnership with the industry-led initiative focused on implementing a harmonized approach to measuring and disclosing emissions, the Partnership for Carbon Accounting Financials (PCAF), to incorporate the PCAF Data Quality Score in its Full GHG Emission Dataset, acting as a market leader to address the importance of data quality. And as the risks associated with greenwashing increase, data quality has rarely played a more vital function in raising the accuracy of reporting standards.

As an initiative born out of the need for greater transparency in the financial system’s carbon transition, PCAF’s global network of 380+ committed financial institutions has supported the development of a methodology to measure the GHG emissions associated with their financial activity. To address the challenge regarding the scarcity of climate data, the PCAF Data Quality Score has been developed to provide a gradient that enables financial institutions that lack high-quality data to begin their decarbonization journey.

Since December 2021, PCAF and CDP have worked in partnership to promote the PCAF Standard and increase the assessment and disclosure of financed emissions amongst financial institutions globally. The collaboration aims to enable transparent disclosures allowing financial institutions and other stakeholders to better understand the climate impact of their portfolios. The assessment and disclosure of financed emissions provide financial institutions with insight into their portfolios’ carbon footprints and alignment with global climate goals.

The partnership was formed with the objective of increasing the number of financial institutions reporting financed emissions in line with the PCAF Standard and – with CDP’s Full GHG Emissions Dataset aligned to PCAF data quality scores, and with improved reporting on the PCAF standards through CDP’s Financial Services Questionnaire – the gap in critical data on climate reporting is beginning to close.

As the need for continued improvement in data quality requires, this partnership will continue to explore opportunities to streamline the reporting of portfolio impact metrics.

The alignment between CDP and PCAF represents their continued commitment to promoting transparency across the financial system and driving consistent climate-related disclosure globally.

# Data quality: Introduction

Data quality is the measurement of how well-suited a data set is to enable financial institutions to calculate the emissions associated with their financial activity to begin the journey to decarbonization. Data quality is therefore fundamental to ensuring that long-term decisions related to decarbonization are both accurate and accountable. The quality of data determines the certainty of a calculation, which has critical implications for all subsequent stages of the decarbonization journey from target setting to implementation.

But beyond the individual financial institutions concerned, there are universal consequences for poor data quality. There is currently a notable discrepancy between the emissions recorded in the atmosphere and the emissions reported by companies. The ability to accurately report emissions is therefore contingent on the continuous improvement of data quality, a process that requires a starting point, even if this means calculations are based on approximate data to begin with. Data quality limitations must not act as a deterrent to decarbonization.

The availability of data varies greatly, requiring financial institutions to rely on approximates that are nonetheless a valuable substitute for a complete data set. When it comes to data quality, 'perfect' is often sought at the expense of 'good' and whilst complete data of the highest quality is ideal, to make the necessary progress towards decarbonization, it is essential to use the highest quality data available for each asset class for calculations and, where relevant, improve the quality of the data over time.

Beyond the availability of data, the lack of standardization when it comes to data quality systems and scoring also impacts the financial system's journey to decarbonization. Across organizations, the concept of data quality has yet to be standardized. This lack of standardization can lead to inaccuracy and inconsistency in emissions reporting between financial institutions, making cross-comparison and benchmarking difficult.

Over the course of this paper, PCAF and CDP set out how their continued collaboration aims to align their respective data quality systems to simplify and streamline the process of measuring and disclosing emissions for the financial system.

# CDP and PCAF data quality scoring

## WHAT ARE THE CDP DATA QUALITY SCORES?

The CDP Data Quality Scores were originally used to indicate at a technical level the accuracy and reliability of data published only in the Full GHG Emissions Dataset.

Data is ranked from 1 to 7, with 7 being the most reliable, (i.e. a company's reported emissions) and 1 being the least reliable (i.e. modelled data with a relatively small sample size). If a company's total emissions data is estimated from more than one type of model, a weighted CDP Data Quality score is calculated based on the revenue activity split of the company.<sup>1</sup>

Since 2022, the current data quality scoring methodology has been aligned to PCAF's data quality scores, as referenced in the Global GHG Accounting & Reporting Standard for the Financial Industry, for listed equity and corporate bonds. As of 2022, CDP is now fully aligned with the PCAF data quality scores as shown in the table below.

## HOW WERE CDP'S DATA QUALITY SCORES DEVELOPED?

CDP recognized the need to portray differing levels of confidence in the data being output in the Full GHG Emissions Dataset.

While reported data is the highest quality and most desirable in terms of accuracy, emissions estimates based on different inputs have varying levels of statistical accuracy. Some industries have well-known direct emissions factors derived from standard production processes, while others must rely on indirect proxy metrics that are another step removed from reality.

The data quality scores were developed to codify this difference. This allows users of the Full GHG Emissions Dataset to understand how accurate the data they are using is but also allows for transparency by identifying companies with low-quality scores and where these scores could be improved.

## PCAF DATA QUALITY SCORING SYSTEM

The PCAF data quality scoring system has been designed to facilitate data transparency whilst encouraging improvements to data quality.

Whilst different asset classes have different data quality scoring options to reflect the nuances of these asset classes, the structure of the data quality scoring system remains the same.

The PCAF data quality score is based on the three data options used to calculate financed emissions.<sup>2</sup> In order of preference, these options are: 1) Reported Emissions, 2) Physical activity-based emissions and 3) Economic activity-based emissions. Options 1 and 2 are preferred over Option 3 from a data quality perspective but financial institutions might use Options 1 or 2 for certain companies and Option 3 for others, in which cases the final data quality score will be determined by an average.

---

<sup>1</sup> Please note that CDP Data Quality scores are not used in either the CDP Scores or in a company's disclosure.

<sup>2</sup> The system used to determine the PCAF Data Quality Score is specific to each asset class and across each asset class, there are minor variations dependent on the type of data available.

These three options are then divided into sub-options depending on the available data, which translate into the five scores of the PCAF data quality scoring system ranked from 1, relating exclusively to verified emissions down to 5.

Whilst a score of 5 reflects relatively low data quality, due to a lack of available data, it represents the necessary starting point for financial institutions. The system in part exists to encourage the continuous improvement of data quality so where the data quality is lower, financial institutions can develop an approach to improve it in time. As data sources become more widely available over time, financial institutions are expected to move up the data hierarchy, thereby improving their data quality score.

## PCAF and CDP alignment

The alignment was built up over a period of a few months between a PCAF and CDP working group and involved matching the different categories of the CDP Data Quality Score (DQS) with the PCAF DQS and tweaking aspects of the CDP modelling methodology to ensure that there was no substantial variation between the two measures. As the CDP DQS is more granular than PCAF, the alignment is one-way (i.e. A PCAF DQS can be derived from a CDP DQS, but not vice-versa).

To determine alignment, the CDP Data Quality Score – once established – is converted to the most appropriate PCAF Data Quality, as shown in the alignment table [See page 7].

This alignment will allow users of the CDP Full GHG Emissions Dataset to utilize it for the purpose of reporting their financed emissions with the added metric of the weighted PCAF Data Quality Score. This will ensure that financial institutions are aware of the potentially significant data quality limitations within their reporting and work with organizations like CDP to support improvements to emissions data gathering.

The next important step to improve the alignment will be CDP assessments of all uploaded documentation supporting a company's third-party verification claim on reported emissions.

CDP and PCAF envision that this alignment embedded within the Full GHG Emissions Dataset will enable improved financed emissions reporting in alignment with PCAF standards and support company engagement by shining a light on poor-quality emissions disclosures that impairs robust financed emissions reporting.

## Conclusion

The continuous improvement of data quality is essential. While the absence of high-quality data should not deter financial institutions from beginning their journey toward decarbonization – greater accuracy enables more precise target setting and a better measure of corporate progress. If financial institutions are to reach a point at which they can accurately measure the emissions associated with their activity, it is essential to strive for better data quality as they continue along their decarbonization pathways.

By aligning the PCAF data quality scoring system with the CDP data quality score, this partnership aims toward harmonizing the concepts around data quality and ensure that PCAF signatories and CDP signatories alike can quickly determine the level of data quality for reporting and disclosure purposes.

There is an ever-growing need for increasingly robust data in this area and the alignment between CDP and PCAF in relation to data quality reflects the need for consensus across the financial system to support financial institutions in measuring their climate impact.

# CDP 2022: Full GHG Emissions Dataset - Version 1.0

## CDP - PCAF Alignment: Data Quality Scoring Guide

CDP Data Quality Scores are used to indicate the accuracy and reliability of data published in the Full GHG Emissions Dataset. Data is ranked from 1 to 7, with 7 being the most reliable, i.e., a company's reported emissions that has passed the Data Analytics team's internal cleaning checks, and 1 being the least reliable. If a company's total emissions data is estimated from more than one type of model, a weighted quality score is calculated based on the revenue activity split of the company.

For 2022, the current data quality scoring methodology has been aligned to PCAF's data quality scores, as referenced in the Global GHG Accounting & Reporting Standard for the Financial Industry, for listed equity and corporate bonds. As of 2022 we are now fully aligned with the PCAF data quality scores as shown below.

CDP Score	PCAF Score	Methodology Type	Description	Example	Alignment Explanation	PCAF Alignment
7	1	Reported emissions	Verified emissions disclosed to CDP.	Reported: CDP (Reviewed)	Fully aligns with PCAF Option 1a ' <b>verified emissions</b> ' of the company are available.	✓
	2	Reported emissions	Verified emissions disclosed to CDP. Verification statement has been reviewed however did not pass quality check.	Reported: CDP (Reviewed)	Fully aligns with PCAF Option 1b ' <b>unverified emissions</b> ' calculated by the company are available.	✓
			Unverified emissions disclosed to CDP; emissions found in company filings.	Reported: Filings (Reviewed)		✓
6	2	Reported emissions	Emissions disclosed to CDP that have been flagged for a units error by the Data Team and multiplied/divided by the correct value.	Reported: CDPdivby1000		✓
		Physical activity-based emissions	Emissions are estimated using primary physical energy consumption data and associated emission factors.	Estimated: Physical Grid Emissions Factor Model (applied to Reported SHEC breakdown)	Aligns with PCAF Option 2a 'emissions are calculated using primary physical activity data of the <b>company's energy consumption</b> and emission factors specific to that primary data'.	✓
5	3	Physical activity-based emissions	Emissions are estimated using physical activity data and associated emission factors.	Estimated: Physical Activity Model (Cement)	Fully aligns with PCAF Option 2b 'emissions are calculated using primary physical activity data of the <b>company's production</b> and emission factors specific to that primary data'.	✓
4	4	Physical activity-based emissions	Emissions are estimated using physical energy consumption data, economic breakdown and associated emission factors.	Estimated: Physical Grid Emissions Factor Model (applied to Reported or Estimated SHEC)	Fully aligns with PCAF Option 3a where ' <b>company's revenue</b> ' is known and 'emission factors for the sector per unit of revenue' are calculated using a multi-variate regression model.	✓
3		Economic activity-based emissions	Emissions are estimated with a multi-variate regression model using revenue and a hybrid climate change focused activity classification.	Estimated: Economic Activity Model (GLM)	Fully aligns with PCAF Option 3a where ' <b>company's revenue</b> ' is known and 'emission factors for the sector per unit of revenue' are calculated using a multi-variate regression model.	✓
2						
1						