WE MEAN
BUSINESS
COALITION

QUALITY MATTERS:
TRANSFORMING ESG DATA FOR BETTER DECISION-MAKING

By Dr. Jane Thostrup Jagd and M.Sc. Sebastian Fischbach
We Mean Business Coalition

CONTENTS

INTRODUCTION

In an era where environmental, social, and governance (ESG) considerations are increasingly integral to business strategy and decision-making, the reliability and usability of ESG data have become paramount. However, significant challenges persist in ensuring its quality, consistency, and accessibility of this critical data. Companies, regulators, data providers, investors, auditors, academics and other users all struggle.

This report, based on the experiences and observations of the We Mean Business Coalition (WMBC), aims to foster a constructive dialogue with these key stakeholders. To maintain a focus on broader systemic improvements rather than specific incidents, we have chosen not to disclose the names of the data providers investigated, companies that have clearly reported incorrect data, or published analyses that clearly have failed to account for poor data quality. Our purpose is to illuminate systemic issues and promote meaningful, collaborative solutions.

We have identified key issues stemming from three primary sources: companies, data providers, and structural factors. For each category, we describe the problem, highlight known upcoming remediation efforts, and propose further actions to address the challenges. We offer concrete advice for changes needed to be made by:

- **Companies:** Identify needed KPIs, establish internal controls, and adopt useful reporting formats.
- **Data providers/investor platforms:** Perform quality checks, identify the required level of detail, and ensure ESG data availability for retail investors.
- **Data users:** Ask critical questions when the data does not "stack up."
- **Systemic changes:** Mandate ESG reporting, assurance, and XBRL tagging; establish more XBRL repositories; improve education (including for national business authorities); and align financial and ESG data boundaries.

Our aim is to raise awareness of these issues and advocate for the establishment of remediation actions that ensure ESG data becomes both reliable for analysis and valid for decision-making. By addressing issues in ESG data quality and usability, and providing practical solutions, we aim to close gaps in reporting, enhance data quality and granularity, and ensure the integrity of the ESG data ecosystem.

Additionally, in the appendix, we provide practical advice for investors, banks, journalists, and academics, as well as data verifiers such as data providers, authorities, and auditors, on identifying and addressing data quality issues when working with ESG data. We hope these stakeholders can benefit from the experiences and approaches we share and, perhaps more importantly, build upon them further.

Through this work, we strive not only to highlight challenges but also to encourage active engagement in improving data collection and verification processes. Ultimately, our goal is to foster an ESG data culture that supports informed and effective decision-making. Together, we can ensure that ESG data becomes a reliable cornerstone for company analysis, sustainable business practices, and investment strategies.

BACKGROUND

WMBC has been a driving force and a founding entity behind several key climate commitments, including the Science Based Targets initiative (SBTi), The Climate Pledge (TCP), and the SME Climate Hub (SMECH), among others. These initiatives are grounded in our theory of change, which posits that if companies commit to reducing their emissions, they are more likely to follow through and achieve those reductions.

However, we acknowledge that we currently lack definitive evidence to support this theory. To address this, we sought to investigate whether companies that make such commitments reduce their emissions more significantly and/or faster than those that do not. Our goal was to explore both correlation – whether companies with commitments reduce emissions faster – and causation, i.e., whether the commitment is a primary driving factor behind the reduction, rather than other variables.

To conduct this analysis, we needed a robust time-series dataset that tracks companies' emissions both before and after their commitment, while also allowing for comparisons with similar companies that have not made such commitments. This is where we encountered a significant challenge—not due to a lack of access to top-tier data providers because we have that in place, but because of the poor quality of historical emissions data provided by these sources.

We have built a comprehensive database of over 18,000 companies that have made one or more commitments—these companies we refer to as WMBC-companies. This database is further supplemented with financial and non-financial data from two of the largest data providers in the world. Additionally, the database includes the same datasets for thousands of companies that have not made any commitments. Finally, it also incorporates information from various NGO repositories.

Despite this extensive dataset, it became evident that the quality of emissions data from as recently as 3-4 years ago is inadequate for meaningful analysis, both for companies that have made commitments and for those that have not. Examples of the issues we encountered include:

- Not all companies with commitments report on emissions. It may sound odd, but it is true.
- Many companies with commitments and reported emissions do not report data further back than 2-3 years. This makes it difficult to identify trends over such a short time span.
- A significant percentage of older data contains blatant errors. While these issues persist, they occur at a much lower rate in recent years. For instance, emissions data from WMBC-companies in 2021 and 2022 had a far lower rate of obvious data issues (0,25%) compared to 2019 (7%).[1] This improvement occurred despite almost twice as many WMBC-companies reporting in 2021 and 2022 compared to 2019. One of the most obvious examples involves companies reporting emissions higher than total global emissions. This is often due to errors on the data provider side. Although such errors are less frequent in more recent years, they still exist.

It is possible to conduct relevant analyses with relatively current data, but the further back you go, the more frequent—and more egregious—these errors become. This pattern suggests a lack of even basic data quality checks in the past. To find examples of logical tests to perform on ESG datasets, please refer to the appendix.

However, many of the challenges we have observed are also being actively addressed in various jurisdictions through new legislation that has either been adopted or is on the verge of implementation, though it has not yet taken full effect.

Hence, rather than attempting to clean and/or adjust the historically flawed data—an approach that could undermine the reliability of our findings—we decided to pause our analysis. For this paper, our focus is therefore to highlight the data quality issues as we have observed them.

We do not claim that this is an exhaustive list of issues, but it represents the challenges we have identified. It is also important to note that, for this analysis, we have primarily concentrated on GHG reporting by companies. Accordingly, many of the examples provided stem from this area. However, many of the issues we observe are similar to those encountered in other ESG areas. We believe this list will therefore be useful to a broader audience. To that end, we will endeavour to highlight where these issues apply more broadly across the ESG sphere.

In the following sections, we describe the ESG data flow, illustrating some of the generic issues within the ESG data landscape. We then provide a detailed analysis of the data quality problems we have encountered, categorizing them based on their sources: companies, data providers, and structural challenges. For each problem, we outline the known upcoming remediations and discuss whether additional measures are needed to fully address the quality issues.

Finally, in the appendix, we provide an overview of the methods we use to assess whether a given dataset we are working with is appropriate and useful for analysis. While we do not claim this to be an exhaustive list of methods and tools, we hope it can serve as an inspirational toolbox for others wishing to incorporate ESG data into their analyses. Perhaps it may even inspire the development of additional data quality inspection tools, which could prove valuable for future ESG data work conducted by companies, auditors, academics, capital providers, and national business authorities.

ESG DATA – HOW TO ACCESS IT?

Currently, there are roughly three types of ESG datasets, and customers can typically choose which to purchase:

1. Raw ESG data: Data reported directly by companies, extracted from their annual reports by the data providers, or collected through surveys sent to the companies.

2. Calculated ESG data: Data artificially estimated based on peer data from similar companies.

3. ESG ratings/rankings: Aggregated scores or rankings based on various elements of ESG data, combining both raw and calculated data.

Most individuals or organizations needing financial and ESG company data for various types of analyses do not extract it manually by reading and copy-pasting from companies' annual reports. Instead, they typically purchase the data from data providers who have consolidated it into large repositories using various methods. However, buying data—especially ESG data—can be quite expensive. Several factors contribute to this cost:

- Identifying the right company reports that contain the required ESG data can be challenging and resource intensive.

- Company reports are produced in as many formats as there are companies, making data extraction both cumbersome and costly.

- Finally, establishing and maintaining repositories that allow users to extract the necessary data for their analyses is itself a resource-intensive task—even after the data has been collected.

In many jurisdictions, it is mandatory to report at least financial data digitally in formats such as XBRL (eXtensible Business Reporting Language) or iXBRL (Inline XBRL).[2] These digital reports are typically accessible—often free or at very low cost—through national business authorities. Consequently, financial data is often available at minimal expense. In contrast, while mandatory XBRL reporting of ESG data is expected to become widespread in many jurisdictions within a few years, it is not yet common. For now, ESG data remains both expensive and difficult to obtain.

Due in large part to the lack of mandatory reporting regulations until now, many companies do not report on ESG metrics. Most often,[3] only large, publicly listed companies in developed countries report—often voluntarily— while privately owned large companies, most SMEs, and many companies in developing countries frequently do not produce any ESG reports. Even when companies do report on ESG metrics, they often provide data for only some components of ESG, and these components can vary widely between companies.

To address these gaps, many data providers have established calculation models to artificially estimate missing data points based on peer companies' data. These models allow the data providers to sell larger and more comprehensive datasets to their customers, such as banks, investors, and academics. However, the quality of these artificially calculated KPIs is often highly imprecise and problematic. For example, Kalesnik, Wilkens, and Zink (2022)[4] conclude: "Their results debunk the belief that third party estimated emissions are a satisfactory substitute for company-reported emissions and call for mandatory and audited carbon emissions disclosure."

Finally, given the lack of comparability in sustainability reports and the challenges of even locating the necessary ESG data, many data providers also sell ESG ratings and rankings. These ratings and rankings often make the work significantly easier for capital providers who need to perform quick ESG analyses of the companies they

wish to invest in, lend money to, or otherwise engage with. To produce these ESG ratings and rankings, data providers still need to collect company data. However, they may also purchase data from other providers, use/ establish calculated data, and often apply various subjective weightings to the data. These subjective weightings allow the data providers to derive the ratings and rankings they sell.

Since WMBC is generally not interested in ratings or rankings, this paper will not delve into the calculation methodologies or transparency issues associated with them. These issues have already been frequently highlighted.[5] However, it is important to note that because raw company data and calculated ESG data serve as the primary inputs for these ratings and rankings, the issues discussed here regarding raw and calculated data will inevitably impact these products as well. Consequently, analyses based on these datasets are also affected.

In the following sections, we will explore each source of the issues in greater detail.

ISSUES STEMMING FROM COMPANY DATA REPORTING

The first set of issues originates from the companies' own reporting habits—before the data providers are even involved.

LACK OF REPORTING ALTOGETHER LEADS TO SKEWED DATA SETS

- **Problem:** As mentioned, many companies fail to report ESG data altogether or report it inconsistently, creating significant challenges for analysts who rely on comprehensive and comparable datasets. Additionally, the companies that do report tend to be skewed towards large, publicly listed entities primarily located in developed regions. Consequently, many analyses are inherently biased due to the underrepresentation of smaller and/or unlisted companies and companies in emerging markets. This skewness will persist until more comprehensive and representative datasets become available.

- **Known upcoming mitigation:** Several legislative frameworks aimed at improving ESG reporting have been adopted or are on the verge of implementation in various countries and regions. Notable examples include the European Union's Corporate Sustainability Reporting Directive (CSRD), the legal adoption of the International Sustainability Standards Board (ISSB) in over 20 countries,[6] the California Climate Disclosure Bill, among others.[7] These regulations are expected to take effect approximately between 2024 and 2028, depending on jurisdiction, company size, ownership structure, and other factors.

- **Further solutions needed:** Broader global adoption of the ISSB standards is necessary to enhance the likelihood of comparable reporting across companies. Expanding the jurisdictional scope of ESG and climate reporting regulations will help ensure better data coverage, leading to more transparent and actionable information for investment and capital allocation. Jurisdictions should also consider including more companies beyond large, listed entities. For instance, large unlisted companies and even listed SMEs could be included to further enhance the breadth and depth of ESG reporting

LACK OF REPORTING ON SPECIFIC KPIs

- **Problem:** Companies often fail to report on specific key performance indicators (KPIs), leading to gaps in essential data, and can make cross-company analyses difficult.

- **Known upcoming mitigation:** Same solution as to "lack of reporting altogether".

- **Further solutions needed:** Same solution as to "further solutions for lack of reporting altogether". But better dialogue with the company's stakeholder may also enhance the reporting of needed KPIs.

INCONSISTENT CALCULATION METHODS

- **Problem:** A lack of uniformity in calculating ESG metrics across companies makes cross-company comparisons difficult. Additionally, companies frequently fail to use consistent calculation methods for their own metrics over time, complicating time-series analyses. This inconsistency also conflicts with basic financial reporting principles,[8] as well as frameworks like the CSRD and ISSB,[9] which require that if an accounting principle is changed or a mistake is discovered, the comparison data must be updated. Similarly, potential targets and plans may also need to be revised to reflect these changes.

- **Known upcoming mitigation:** Same solution as to "lack of reporting altogether".

- **Further solutions needed:** Same solution as to "Further solutions for lack of reporting altogether". However, education initiatives and more widespread mandatory auditing/assurance of ESG reports will further enhance accuracy and reliability.

  The most important step, however, is for companies to establish better internal controls.[10] Robust internal control environments should include comprehensive data manuals that clearly define how various KPIs are calculated. This would lead to much better and more consistent reporting from companies. In turn, companies would benefit from faster, better, and more cost-effective assurance processes.

MISUNDERSTANDING REPORTING REQUIREMENTS

- **Problem:** Some companies misunderstand ESG reporting requirements, which can lead to inaccuracies. For example, some companies report only carbon dioxide (CO_2) emissions instead of carbon dioxide equivalents (CO2e), resulting in underreporting. This is problematic because the Global Warming Potential (GWP) factors of the seven Kyoto Protocol gases vary significantly, with all six non-CO_2 gases having higher GWP factors than CO_2. Additionally, some companies fail to include all seven Kyoto gases in their equivalent conversions.

  Another example could be that many companies do not segregate their greenhouse gas (GHG) emissions across the three scopes, instead reporting only a single aggregated GHG figure.

  Similarly, some companies fail to distinguish between location-based and market-based Scope 2 emissions, instead labelling them simply as "Scope 2." This lack of granularity is problematic, as most robust analyses require such distinctions.

  A fourth common misunderstanding occurs when companies equate energy consumption with electricity. This is incorrect, as energy consumption includes the total energy from all Scope 1 and Scope 2 sources, plus renewable energy, while electricity is merely a subset of Scope 2.

  But many more such misunderstandings are evident across companies.

- **Known upcoming mitigation:** Same solution as to "lack of reporting altogether".

- **Further solutions needed:** Same solution as to "further solutions for inconsistent calculation methods". However, additional steps should include encouraging data users, such as investors, banks and other stakeholders, to raise questions during companies' AGMs or capital market days when reported data does not "stack up." Challenging companies with their data quality will increase their awareness of the need for accurate, reliable and useful data.

PRESENTATION OF DATA IN NON-USER-FRIENDLY FORMATS

- **Problem:** Traditionally, companies have often used sustainability statements for communication purposes—sometimes even as part of their marketing efforts. As a result, data is still frequently presented in beautiful, colourful illustrations or narrative text. While this approach may serve marketing objectives, it is counterproductive now that ESG data is being integrated into financial reports and used by capital providers in their analyses. Such presentations make it difficult to identify, extract, and effectively use the data.

- **Known upcoming mitigation:** The adoption of mandatory XBRL reporting for ESG data in many jurisdictions will require digital tagging of data, facilitating easier data extraction into central repositories.

- **Further solutions needed:** More jurisdictions should mandate XBRL reporting, as most already do for financial reports. However, we recognize that XBRL tagging requires discipline to be effective. Therefore, we recommend mandatory assurance of XBRL-tagged reports. Authorities in jurisdictions that have adopted or will adopt XBRL reporting could perhaps also benefit from studying India's[11] experience, where mandatory XBRL tagging of ESG reports was implemented for the 1,000 largest listed companies starting in the 2022/23 reporting year.

Additionally, companies should consider how to present their data to minimize the risk of misinterpretation. Tabular reporting for ESG metrics simplifies data access for investors and analysts, reducing misreading and extraction errors. It would also make XBRL tagging significantly easier for the companies themselves.

ISSUES STEMMING FROM DATA PROVIDERS

The following issues arise from the practices of data providers:

DELAYED DATA AVAILABILITY

- **Problem:** ESG data provided by data providers is often outdated by 1.5 years or more. Many companies now report their ESG data simultaneously with their financial data, often in integrated reports, typically 1–5 months after the reporting year. While this timing is reasonable, the extraction of non-financial data by data providers is very slow. As a result, even the largest data providers take an additional year to finalize the 'new data.' For instance, company data reported for the 2022 reporting year was only fully available by summer 2024.

- **Known upcoming mitigation:** The use of XBRL tagging for financial and ESG reporting is expected to reduce this lag. This improvement will become evident once large public repositories, such as the EU's European Single Access Point (ESAP - expected in 2027)[12] or UN's GFANZ-related Net Zero Data Public Utility (NZDPU - 2025,[13] though initially limited to historical data from CDP), become operational. These repositories will provide faster, cheaper and easier access to ESG data for both data providers and users, including capital providers, NGOs, and academics.

- **Further solutions needed:** More jurisdictions should mandate XBRL reporting and establish publicly accessible repositories for the XBRL files.

DATA EXTRACTION METHODS LEADING TO ERRORS

- **Problem:** Data providers in the market use various methods to extract or collect data for their databases. Most rely on manual extraction performed by humans, some use machine reading, others send surveys to companies, and some employ a combination of these approaches.

  Manual extraction and survey responses often result in errors, such as unit mistakes, typos, and missing data points which were buried in text or illustrations. Databases relying on machine reading face challenges as well, as this method is not always mature or precise. Machines can also make unit mistakes, miss data points, and struggle to interpret illustrations and graphs. However, while errors from machine reading tend to be systematic, errors from manual extraction and surveys are typically more random.

- **Known upcoming mitigation:** The adoption of XBRL tagging will significantly reduce the need for these flawed data extraction and collection methods. This advancement could also negate the need for raw ESG data to be sold, as all organizations and individuals would be able to extract data directly from central repositories for free.

- **Further solutions needed:** More jurisdictions should mandate XBRL reporting – but do also see 'Further solutions needed for quality checks'.

LACK OF QUALITY CHECKS AND FAILURE TO UPDATE HISTORICAL DATA

- **Problem:** We would argue that data providers, at least historically, appear to have implemented only limited quality checks. As indicated in the introduction, we have observed cases where individual company emissions exceed total global emissions—a clear impossibility. When we cross-checked these figures against the company's own reports, it became evident that the data providers' data did not match the company's actual disclosures. This indicates an extraction error.

  In all fairness, we have also noted improvements in data quality verification processes over the past few years. As previously mentioned, more recent data contains far fewer obvious errors compared to older datasets. However, historical data remains significantly flawed and uncorrected, making time-series analyses challenging.

- **Foreseeable Problem:** With the introduction of new mandatory standards, reporting companies may find themselves needing to change accounting methodologies for certain datasets they have previously reported. It remains unclear whether this will result in the erasure of historical data (to simplify assurance processes) or in restatements of historical data. Likely, both approaches will be used, as both are permitted/required under standards like ESRS and ISSB, as described previously. Potential concern arises from the current practice of some data providers, who do not update historical data. It is foreseeable that some providers may fail to recognize the need to extensively restate or erase historical data that is no longer valid and hence comparable. If this lack of action persists, it will hinder the reliability of time-series analyses.

- **Known upcoming mitigation:** XBRL tagging is expected to reduce manual extraction and associated errors, leading to more error-free ESG data repositories.

- **Further solutions needed:** More jurisdictions should mandate XBRL reporting and the assurance of XBRL files, as this will enhance the quality of the data within them.

  XBRL taxonomies should include stronger internal controls to ensure the validity of the XBRL files themselves.

  Better training for companies on how to use XBRL tools effectively will help reduce tagging errors. Experiences from financial XBRL reporting, as well as insights from India's implementation of ESG XBRL-tagging, demonstrate that digital tagging does not entirely prevent mistakes, as the tagging process remains partly manual.

  Furthermore, data providers should implement more robust quality checks on the data they extract, ensuring accuracy over time and across both financial and non-financial datasets.

LACK OF GRANULARITY IN DATA

- **Problem:** ESG data providers often fail to provide sufficient detail, such as distinguishing between market-based and location-based Scope 2 emissions, segregating renewable and non-renewable energy consumption, offering taxonomy data detailed down to activity numbers or even to verify basic information, such as whether specific data points have been assured or audited. Sometimes this lack of granularity is due to the mentioned misunderstandings of requirements at company level – but often not. This lack of granularity makes it difficult to conduct analyses.

- **Known upcoming mitigation:** XBRL tagging has the potential to enable more granular data capture, which is expected to improve as the technology is further refined. However, there is a risk that some companies might use CLOB (Character Large Object) tagging, a type of block tagging for text, which could fail to deliver the necessary level of detail.

- **Further solutions needed:** Greater collaboration between data users and providers could help ensure that repositories capture the required level of detail. Additionally, clearer rules for XBRL tagging should be established to ensure that the resulting XBRL files are practical and meet users' needs.

STRUCTURAL ISSUES

The following issues are more structural in nature.

LACK OF A CENTRAL REPOSITORY

- **Problem:** There is currently no central repository for ESG data, making it difficult for stakeholders and analysts to access information without incurring significant costs.

- **Known upcoming mitigation:** The previously mentioned ESAP and NZDPU initiatives aim to make ESG data more accessible and are expected to become operational within the next few years. Additionally, national business authorities are expected to host repositories for ESG data from companies, similar to how financial data is managed today. Interesting pioneering solutions are already in place for instance in India, Japan[14] or Singapore.[15]

- **Further solutions needed:** More data should be gathered to ensure accessibility, including for non-listed companies and SMEs, which are often overlooked by current data providers due to limited demand from their primary customers, who often focus on large, listed companies in developed markets.

Investor platforms should offer raw data (e.g., Scope 1 and Scope 2 GHG emissions) alongside financial KPIs, enabling retail investors to access critical climate information without relying on expensive data providers. In this way, investors, analysts and others can, for instance, conduct integrated analyses of companies' price-earnings ratios versus their GHG emissions per share. This allows them to compare the price of the most profitable companies with their green risk profiles. Additionally, investors can identify companies that are "on their way"—not fully green yet—and uncover "hidden gems" with alpha potential.

Integrating ESG data feeds from companies into publicly available repositories and trading platforms would significantly improve the timeliness and accessibility of data.

LACK OF VALIDATION BY NATIONAL AUTHORITIES

- **Problem:** Currently, national business authorities manually and/or automatically scrutinize financial reports for obvious violations of financial reporting legislation. However, with ESG reporting regulations—including mandatory reporting in some jurisdictions—only now coming into effect, national authorities have yet to take on the challenge of scrutinizing ESG reports for clear violations. So far, legal cases have primarily focused on violations of greenwashing regulations[16] particularly by financial institutions in their fund naming practices. As Wang, Sun & Miao (2023)[17] observe: "Currently, few studies focus on the behaviour and mechanism of corporate carbon emission data fraud."

- **Known upcoming mitigation:** Certain jurisdictions (e.g., the EU, Australia, Singapore) are beginning to mandate third-party assurance of ESG data. This requirement will help address the worst mistakes and violations.[18] However, it does not necessarily mean national authorities will begin supervising ESG reporting. That said, organizations like ESMA have taken initial steps, as evidenced by their inclusion of sustainability statements in the European Common Enforcement Priorities (ECEP) for 2024.[19]

- **Further solutions needed:** Expanding mandatory assurance requirements across jurisdictions will improve the reliability of ESG data. XBRL taxonomies should incorporate more automated validation controls to detect anomalies, and assurance of XBRL reporting must become a natural next step.

  While auditors are currently upskilling to meet the demands of ESG reporting, it would also be beneficial for employees of national business authorities to undergo similar training to enhance their ability to oversee ESG reporting incl. the XBRL reporting effectively.

MISALIGNMENT OF GHG AND FINANCIAL REPORTING BOUNDARIES

- **Problem:** GHG reporting boundaries do not always align with financial reporting boundaries, complicating integrated analyses with financial data such as revenue or cash flow. This misalignment also creates challenges for cross-company analyses. For instance, GHG data may sometimes cover more, and other times less, activity than the financial data—this is true even when considering only the companies' own fuel combustion and use of electricity and district heating (GHG Scope 1 and 2).

  Both the EU's ESRS E1 and ISSB's IFRS S2 have adopted "operational boundaries" for GHG Scope 1 and 2 reporting. However, neither standard defines these boundaries explicitly; instead, they refer to the GHG Protocol from 2004, a standard created before the issuance of IFRS 10 (Consolidated Financial Statements) and 11 (Joint Arrangements) or the updates to US GAAP ASC 810 on financial boundaries.[20]

  While it is commendable that the European Commission (EC) and ISSB have aligned on the concept of "operational boundaries," it remains an unfortunate solution. This misalignment with the financial reporting boundaries hinders logical analyses, such as comparisons of emissions against a company's activity levels, revenue, or cash flow. Furthermore, it is not future proof, as it fails to account for potential updates to financial standards.

- **Known upcoming mitigation:** None – unfortunately.

- **Further solutions needed:** It would be essential for the EC and ISSB, perhaps in collaboration with the GHG Protocol, to acknowledge this issue and rectify it by aligning GHG reporting boundaries with financial reporting boundaries. The GHG Protocol should avoid attempting to define their own boundaries and instead defer to existing financial standards, ensuring future-proof comparability.

  Such an adjustment would not reduce reported emissions; for instance, emissions from operated but non-owned parts of a joint operation or from leased-out assets with personnel onboard would still be counted—but as Scope 3 emissions. However, this change would make comparisons of emissions against company activities and profits, as well as cross-company analyses, far more logical and useful.

CONCLUSION

Our analysis of ESG data quality has revealed critical challenges that must be addressed to ensure reliable and actionable insights for stakeholders. As highlighted in this report, issues stem from multiple sources, including company reporting practices, data provider methodologies, and structural shortcomings in the broader ESG data landscape. These issues complicate cross-company comparisons, time-series analyses, and integrated assessments that combine ESG and financial metrics.

KEY TAKEAWAYS:

1. **Data challenges across the ESG spectrum:**
   While raw ESG data holds the greatest potential for robust analysis, its availability is still limited, and its quality is often inadequate. Calculated data and ESG ratings/rankings, while more accessible, introduce additional concerns, including lack of transparency and inconsistent methodologies.

2. **Improvements in regulatory frameworks:**
   New regulations such as the EU's CSRD and similar initiatives globally such as the ISSB promise better data quality and granularity. However, their success depends on widespread adoption, clear guidelines, and rigorous enforcement.

3. **Role of data providers:**
   Data providers play a critical role in consolidating and disseminating ESG data but must address significant delays, inconsistencies, and inadequate validation processes. Enhanced quality checks, better training, and adherence to standardized practices will be essential for improvement.

4. **Structural reforms needed:**
   The establishment of central repositories for ESG data and better integration of financial and ESG reporting boundaries are necessary steps. Aligning GHG reporting with financial boundaries, for example, would enable more logical and comparable analyses.

5. **A call for collaboration:**
   Achieving meaningful improvements requires collaboration among regulators, data providers, companies, and stakeholders. This includes adopting shared standards, providing training, and integrating robust internal controls at the company level.

CHANGES NEEDED FOR COMPANIES:

- Companies should ensure better dialogue with stakeholders to identify the necessary KPIs.

- Companies should establish stronger internal controls. A robust internal control environment must include comprehensive data manuals that clearly define how various KPIs are calculated.

- Companies should recognize that tabular reporting of ESG metrics simplifies data access for investors and analysts, reducing misreading and extraction errors. This approach would also make XBRL tagging significantly easier for the companies themselves.

CHANGES NEEDED FOR DATA PROVIDERS/INVESTOR PLATFORMS:

- Data providers should implement more robust quality checks on the data they extract, ensuring accuracy and reliability across time and datasets.

- Data providers should foster improved dialogue between data users and providers to ensure repositories capture the required level of detail.

- Investor platforms should offer raw ESG data (e.g., Scope 1 and Scope 2 GHG emissions) alongside financial KPIs. This enables all investors, including retail investors, to access critical climate information without relying on expensive data providers. Including such data would also significantly improve the timeliness and accessibility of ESG data from both companies and data providers.

CHANGES NEEDED FOR DATA USERS:

- Data users, such as investors, banks and other stakeholders, should raise questions during companies' AGMs or capital market days when reported data does not "stack up." Challenging companies on their data quality will increase their awareness of the need for accurate, reliable, and useful data.

SYSTEMIC CHANGES NEEDED:

- Broader global adoption of ISSB standards is necessary to enhance the likelihood of comparable reporting across companies. Jurisdictions should also consider including more companies beyond large, listed entities.

- ESG reporting legislation should mandate assurance of ESG reports, as this further enhances accuracy, reliability, and hence, usability.

- More jurisdictions should mandate XBRL reporting and assurance of XBRL-tagged reports.

- Jurisdictions or regions should establish publicly accessible repositories for XBRL files.

- These XBRL repositories should also include data from non-listed large companies and SMEs, often overlooked by commercial data providers due to limited demand from their primary customers, who tend to focus on large, listed companies in developed markets.

- XBRL taxonomies should include stronger internal controls to ensure the validity of the XBRL files. Clearer rules for XBRL tagging should also be established to ensure that the resulting XBRL files are practical and meet users' needs.

- Education initiatives for ESG reporting, ESG auditing/assurance, and XBRL tagging should be included in regular business school curricula alongside financial reporting and auditing classes.

- While auditors are currently upskilling to meet the demands of ESG reporting, it would be equally beneficial for employees of national business authorities to undergo similar training to enhance their ability to oversee ESG reporting effectively – including validation of XBRL tagging.

- The European Commission and ISSB, perhaps in collaboration with the GHG Protocol, should acknowledge the boundary issue and work to align GHG reporting boundaries with financial reporting boundaries.

FINAL REFLECTION:

While current and historical ESG data quality poses significant challenges, progress is evident, and solutions are within reach. As new regulations and technologies take hold, the availability and reliability of ESG data will improve. Until then, stakeholders must navigate the complexities with caution, focusing on incrementally enhancing data quality while advocating for systemic changes. By addressing these issues collaboratively, we can ensure that ESG data fulfills its potential as a cornerstone for analyzing companies' performance and prospects.

Our purpose with this paper was to illuminate systemic issues, promote meaningful and collaborative solutions to ESG data quality challenges, and enhance the dialogue on the solutions needed from various stakeholders. We hope to have achieved this.

APPENDIX: ADVICE ON HOW TO IDENTIFY DATA QUALITY ISSUES

Should you find yourself in a position where you would like to use ESG data for your analyses but are unsure whether the data is of suitable quality, we would like to offer a set of recommendations. This list is by no means exhaustive; rather, it represents methods we have identified and found useful. We hope that others—including investors, banks, journalists, and academics, as well as data verifiers such as data providers, authorities, and auditors—can benefit from these approaches and perhaps even build upon them further.

SEGREGATE BETWEEN DATA TYPES – RAW DATA, CALCULATED DATA OR RATINGS/RANKINGS

As previously indicated, there are at least three distinct types of ESG data; raw data, calculated data or ratings/rankings. Carefully consider the type of analysis you intend to conduct and, consequently, which data type will be most appropriate.

While we await the full implementation of new regulations, the rarest and most difficult data to access remains the data directly reported by companies (raw data). Limiting your analysis to raw data may reduce the volume of available information. However, using only company-reported data might be crucial for ensuring quality and reliability. Most large data providers do offer the option to segregate between reported and calculated data, which can help in this regard.

If having a large dataset is essential, you may choose to include artificially calculated data, possibly as a supplement. However, as noted earlier, this type of data should be treated with caution, as it is often based on data from the company's peers and is likely to be imprecise - sometimes significantly so. The use of artificially calculated data can to some extent be compared to the statistical method of "bootstrapping" . Both methods rely on the assumption that the collected data is representative of the entire population, allowing the sample to be enlarged based on this assumption. However, there is at least one significant difference: Bootstrapping is a transparent method, whereas artificially calculated data often relies on external models and assumptions that may not be entirely clear to the direct data user or to subsequent users of the analysis results. Overall, one should exercise caution with artificially calculated data, using it perhaps only for statistical analyses of entire populations—and certainly not for individual company analyses, such as those for investments or lending. It should, of course, also be clearly stated in the methodology that artificially calculated data has been used.

You may also need to evaluate ESG ratings or rankings. Sometimes, ratings or rankings are used as substitutes for raw or calculated data, possibly because they are more readily available. In some cases, data providers do not allow users to extract the underlying data behind their ratings or rankings. As discussed previously, these ratings and rankings are known to have significant calculation issues, a lack of transparency, and subjective interpretation concerns. Moreover, they may not fully align with the specific aspects you wish to investigate. Ratings and rankings are also not assured, whereas raw ESG data—especially in the future—is more likely to be audited/assured.

In summary, carefully consider which type of ESG data is most suitable for your analysis.

CHOOSE YOUR DATA PROVIDER CAREFULLY

There are many ESG data providers in the market, varying in size, cost, and focus. Some are large, comprehensive providers, while others are smaller and more specialized. Some are very expensive, while others are more affordable. Certain providers have extensive metadata sets they collect data against, while others focus on narrower, specialized areas. Some even create their own KPIs, which may or may not suit your needs. Additionally, some providers only offer ratings/rankings without access to the raw company data behind them, which limit transparency. Consider also, that while some providers offer robust technical download solutions, such as APIs for seamless data downloads, others do not.

When choosing a data provider for your project, consider the following questions. If possible, ask the provider for a test period to evaluate whether their technical setup meets your needs:

- **Budget:** Do we have the budget to purchase data, or should it be free (or nearly free)? If opting for free data, you may need to compromise on the scale and/or quality of your analysis.

- **Technical requirements:** Does the data provider offer the technical download solutions we need?

- **Timeliness:** Is the data updated with the frequency and speed we require? Often, ESG data is not updated as quickly as desired, which comes as a surprise to many.

- **Geographic or sector-specific needs:** Do we need global data, or is our focus on a specific geographic region or sector? And if global data are needed, does the data provider truly offer global coverage, or is it skewed?

- **Relevant KPIs:** Does the data provider offer the KPIs we need? Some providers specialize in specific topics, which can be useful but may also cause confusion if definitions differ between the provider and your needs. Additionally, ESG providers often lack comprehensive financial data (e.g., revenue or market value), which might be essential for integrated analyses. In such cases, combining datasets from multiple providers may be the best solution. (See the note below on identifying primary and supplementary data providers.)

- **Data scale vs. quality:** Is it necessary to include thousands of companies in your dataset, or could your analysis be conducted with a smaller sample (e.g., 100 companies)? If the latter, manual data collection may yield higher quality results and faster data access.
    - In our experience, manually extracting data from a full report typically takes around two hours per report (some more, some less, especially for a limited number of KPIs).
    - Alternatively, consider using machine reading or AI for data extraction if you have access to a well-organized repository of report files and a well-constructed prompt. For smaller or more detailed analyses, manual data collection may provide better results.

- **Master Data Discipline:** Some data providers excel in ensuring clean, disciplined master data for company identification, particularly expensive data providers favoured by financial institutions. However, data from NGOs, while often insightful, may lack such rigor. We often encounter the following issues with NGO databases, which typically stem from their reliance on company names as key identifiers instead of using official company IDs, such as ISIN (International Securities Identification Number), LEI (Legal Entity Identifier), and/or Company PermID (Permanent Identifiers), among others. Therefore, it is crucial to test the database thoroughly before using it.
    - **Duplicates:** Company names may be spelled differently (e.g., "Limited" vs. "Ltd."), or the same company may appear under both its legal name and brand name. Ensure duplicates are filtered out.
    - **Local scripts:** Company names may be in local languages or scripts, complicating integration with other datasets. Consider filtering out or re-naming.

- **Parent-subsidiary confusion:** Both parent companies and their subsidiaries may be listed as individual entities, inflating the database size. Most analyses should only focus on consolidated group data, so filter out subsidiaries—if you can identify them (which isn't always straightforward).

- **Non-companies:** Some entries may represent organizations that are not companies. Ensure these are filtered out.

- **Outdated names:** Company names often change due to mergers, rebranding, or other organizational changes. Databases that lack governance structures for managing such changes may be difficult to use, especially for time-series analyses. Check the database's governance structure to ensure it can handle these changes.

IDENTIFY PRIMARY AND SUPPLEMENTARY DATA PROVIDERS

You may need to combine data from multiple data providers, and in some cases, these providers will offer data on the same KPIs. In such situations, it is advisable to have a clear strategy for determining which provider will serve as the primary data source and which will be supplementary. Here is how we approach it:

- Extract the data for the shared KPIs from each provider and compare them. Sometimes, the values will not match, especially for data from just a few years ago.

- Test, for example, the 50 largest discrepancies by cross-checking them against the companies' own reports. This process will help you identify which data provider offers the highest-quality data for a specific KPI and which provider should only be used as a supplementary source.

Such strategies should, of course, be included in your methodology description.

MAKE OVERALL LOGICAL ANALYSES OF THE DATA QUALITY

- **Outliers:** To evaluate data quality, start by identifying outliers—unusually large or small values that may appear illogical perhaps given the company's sector. Investigate whether these values align with what the companies reported themselves and whether there are plausible explanations for the anomalies.

- **Illogical datasets:** Look for unexpected data combinations that might indicate potential errors. For example, a company reporting Full-Time Employees (FTEs) but no Scope 2 GHG emissions is unusual, as the presence of employees typically implies electricity usage. While there may be explanations (e.g., the company generates its own electricity through a generator (Scope 1) or a windmill (no emissions), it could also indicate issues like misreported/miscollected Scope 2 data. For instance, the company may have purchased green certificates, which should only reduce the market-based Scope 2 emissions, while the data provider failed to segregate location-based and market-based Scope 2 data. Either way, you must investigate and understand the cause of such illogical combinations.

- **Data development:** Examine trends in the data over time and flag anomalies—whether unusually small or large changes. Cross-check these discrepancies with the companies' reports, which may offer credible explanations for the variations. However, you might also uncover errors such as typos, unit mistakes, or misreadings from the data provider. In some cases, you may find that a company has changed its accounting policies, but the data provider did not update historical data to maintain comparability over time.

- **Normalize the data:** Validate and assess the completeness of the data by comparing it with contextual information, such as financial metrics (e.g., revenue or production quantities). This process can help you create "normalized" data that accounts for differences in activity levels over time and/or across peers. If the normalized data appears inconsistent compared to peers or time trends, investigate further by referring to the companies' own reports. While there may be valid explanations, such anomalies often signal subpar data quality.

- **Be aware of re-organization:** Companies frequently undergo mergers, acquisitions, or spinoffs. These reorganizations generally do not affect historical financial data, hence ESG data should be treated similarly to maintain the integrity of integrated analyses.

  Normalized ESG data reflecting company activity should still make sense, but these reorganizations can complicate time-series analyses. Additionally, company targets and plans will often need to be restated following such changes. For analysts interested in time-series analyses, it is essential to identify companies with significant reorganizations, as these can introduce noise into the analysis.

CONCLUDE WITH QUALITY TESTS

If the data quality is significantly subpar, consider whether a better data source is available. You may be able to rectify incorrect data, especially if only a small percentage of the records are flawed. Alternatively, you could redefine the scope of your analysis to work with the data as is. However, if the quality issues are extensive, you might reach the same conclusion we have in some cases: to wait until better-quality data becomes available.

ENDNOTES:

1. Keep in mind that this only highlights the most obviously poor-quality reports when viewing the records in the sample as a whole. Beyond this, there are likely to be many more poor-quality reports that would require a closer examination of individual records to identify.

2. See more here: eXtensible Business Reporting Language and for iXBRL Inline XBRL. iXBRL is a report format that combines both human-readable and machine-readable data in a single document. While XBRL can be challenging for humans to interpret, iXBRL is often a more user-friendly solution.

3. See also Mou & Ma (2023), A Study on the Quality and Determinants of Climate Information Disclosure of A-Share-Listed Banks, Sustainability 15(10) 8072, Full text.

   Janicka & Sajnóg (2023) The Quality of Environmental Data Disclosure. The case of Public Companies from the European Union Countries, European Journal of Sustainable Development, Vol 12, no 4, Full text

4. Kalesnik, Wilkens & Zink (2022), Do Corporate Carbon Emissions Data Enable Investors to Mitigate Climate Change?, The Journal of Portfolio Management Novel Risks 2022, 48 (10) 119 - 147, Full text

   But do also see: Bajic, Kiesel, & Hellmich (2023), Handle with Care: Challenges and Opportunities of using Company-Level Emissions Data for Assessing Financial Risks from Climate Change, Journal of Climate Finance Vol 5, Full text

   And then read this pre-print paper: Caprioli, Foschi & Crupi (2024) Denoising ESG: Quantifying Data Uncertainty from Missing Data with Machine Learning and Prediction Intervals, Full text

5. See Berg, Kölbel, Rigobón (2022), Aggregate Confusion: The Divergence of ESG Rating, Review of Finance, Volume 26, Issue 6, November 2022, Pages 1315–1344, Full text

   But do also read this famous piece, which has been published many times – but this is the latest version from 2023: The Sustainability Institute by ERM (2023) Rate the Raters, Full text

6. As of 18.11.2024, the list of jurisdictions who fully or partially have or will adopt ISSB is as follows: Hong Kong, South Africa, Singapore, UK, Philippines, Pakistan, Türkiye, Australia, Nigeria, Malaysia, Singapore, India, Canada, Brazil, China, Japan, Uganda, South Korea, Chile, Switzerland, and Mexico. See updates on this site: IFRS - Jurisdictional sustainability consultations

7. PwC (2024) Sustainability reporting adoption tracker – last updated 1 October 2024, PwC: Sustainability reporting adoption tracker

8. See for instance IAS 1, no 38 and IAS 8, no 22

9. See ESRS 1 (7.1, 7.4 and 7.5) and also ESRS 2 (13+14) – or see IFRS S1 (70, 71,83-86 and Appendix B: B40-B45)

10. See more here: IFAC, WMBC & GAA (2024) Building Trust in Sustainability Reporting and Preparing for Assurance: Governance and Controls for Sustainability Information, Full text

11. XBRL & At Quest (2024) Unearthing Insights from India's ESG Disclosures – from BRSR for Financial Year 2022-23, Unearthing Insights from India's ESG Disclosures

12. See more here: The ESAs finalise rules to facilitate access to financial and sustainability information on the ESAP - EIOPA

13. See more here: CDP and the Net-Zero Data Public Utility expand partnership to power the net-zero transition - CDP

14. The Japan Exchange Group (JPX) offers an ESG Knowledge Hub, providing information and guidance on ESG disclosure practices. See more here: https://www.jpx.co.jp/english/corporate/sustainability/esgknowledgehub/index.html?utm_source=chatgpt.com and here: Application of XBRL to Timely Disclosure Information | Japan Exchange Group

15. See more about the interesting Greenprint Project in Singapore, which is more than a XBRL reporting project and also works with potential blockchain solutions etc.: Green FinTech

16. Consider, for instance, the cases of DWS Group (formerly part of Deutsche Bank), Goldman Sachs, Vanguard, Mercer and Active Super —among others. What they all have in common is that they misled investors by using fund names or suggesting ESG criteria that the funds did not actually meet.

17. Wang, Sun & Miao (2023), Management of enterprise carbon emissions data falsification considering government regulation and media monitoring, Frontiers in Environmental Science Vol 11, Full text

18. Gipper, Sequeira & Shi (2024) Carbon Accounting Quality: Measurement and the Role of Assurance, Stanford Business Working paper no 4186, Full text

   But also read: Berg, Huidobro & Rigobon (2024) On the Importance of Assurance in Carbon Accounting, Full text

19. See more here: ESMA32-193237008-8369 European common enforcement priorities for 2024 corporate reporting

20. IFRS 10 (Consolidated Financial Statements) replaced IAS 27 (Consolidated and Separate Financial Statements) and SIC-12 (Consolidation – Special Purpose Entities). IFRS 11 (Joint Arrangements) replaced IAS 31 (Interests in Joint Ventures) and SIC-13 (Jointly Controlled Entities – Non-Monetary Contributions by Venturers) – both IFRS 10 and 11 were issued in 2011 and took effect from 2013. Especially IFRS 11 is important when we talk about "operational boundaries" – but do also read IFRS 16 on Leasing, which was issued in 2016 with effect from 2019.

   ASC 810 (Consolidation) from US GAAP was issued in 1959, but has been updated frequently, for instance in 2003, 2007, 2009 and 2015. See also ASC 842 (Leases) which was released in 2016 and took effect for listed companies from 2019 and private companies from 2022, and has been partially updated since release. ASC 842 aligns conceptually with IFRS 16 in requiring on-balance-sheet recognition for most leases but retains certain distinctions, such as the dual lease classification (operating vs. finance leases).

21. Davidson & Hinkley (1997) Bootstrap Methods and Their Applications, online publication available from Cambridge Series in Statistical and Probabilistic Mathematics (2013), Full text