

ESG data lineage: Why traceability matters more than metrics

The broad narrative around ESG reporting generally begins with talk of metrics. Organizations often engage executive time in conceptualising metrics that would help measure Scope 3 emissions, various diversity ratios, or even quantify renewable energy usage. However, regulators and industry experts are beginning to focus on a different nuance. It is how traceable the ESG data actually is to its source. Looked at differently, the underlying question is: Is the data lineage defensible at the time of assessment? This is the most crucial area of consideration for management while setting out their data model, processes, and systems. Without a robust data lineage, ESG metrics, even if well thought out, are not going to be able to stand the rigour of an assessment.

It is important to consider that the general trend for sustainability reporting is to move away from self-disclosed metrics to more regulated, assurance-ready reporting. Frameworks such as the International Sustainability Standards Board (ISSB) and the European Sustainability Reporting Standards (ESRS) stipulate that businesses show working evidence that the sustainability metrics are accurate, traceable, reliable, and reproducible.

Understanding ESG data lineage

Considering data lineage requires analyzing the end-to-end lifecycle of each data point: its origin, business logic applied, aggregation with other data, controls imposed, and the final reporting destination. For instance, a Scope 3 emissions metric may originate from shop floor meter readings, merge with ERP system consumption data, undergo emission factor calculations in software, and finally be consolidated at the group reporting level for disclosure. Each process step introduces possible errors or assumptions. Data lineage ensures data is traceable, robust, and defensible.

Why lineage matters more than metrics?

A wide variety of organizations are found to lay great emphasis on choosing the most appropriate metrics. However, these metrics often fail to be reconstructed during and assessment. When regulators or assurance agencies look at the metrics, they usually ask questions around the point of origin or the number of transformations or even the underlying methodology which resulted in the final metric being reported. Organizations who are unable to answer one or more of these queries typically lose credibility and are adversely impacted during the assessments.

How regulators expect traceability to be ensured

The UK financial reporting council in its 2023 thematic review on climate related metrics observed that a lot of organizations demonstrated the ability to review and report underlying climate data but when tested further fails to demonstrate the end-to-end traceability of the information. The FRCS report highlighted particular areas of weakness such as:

- Ambiguous ownership of the source data
- Adjustments and assumptions that were not clearly documented or approved
- Inappropriate mechanisms to trace the reported metrics back to the source data

The European Securities and Markets Authority prioritises consistency and traceability of sustainability information, particularly in areas where businesses are found to be changing methodologies or assumptions across reporting time frames. The authorities have repeatedly set out directives asking the organizations to be able to explain how the data has flowed through their operational systems right up to the point of public disclosure.

On the other hand, the ISSB standards drive connectivity between underlying data systems and internal control mechanisms. It expects that the ESG disclosure should have been subjected to appropriate governance standards that are commensurate with those of financial reporting. This includes a clearly audible linkage from the source to the point of reporting and regular governance oversight of the same. As such, the quality of ESG reports is no longer judged by just the sophistication of the metric reported but by how traceable the underlying information actually is.

Change logs and their criticality to ESG data systems

In any reporting scheme, the metrics evolve through the reporting life cycle. They typically are adjusted for revisions in emission factors, updates to storage data, improvements to the methodology, or even changes to system versions. A change log is expected to capture precisely this information. It helps to answer:

- What information has materially changed?
- At which points of time it changed?
- Who approved the change?
- Why was the change made?

Without adequate change logs, organizations will find it difficult to explain how the final metric has evolved.

It is typically seen that corporate climate disclosures change across different reporting years because of evolving methodologies. For instance as organizations adopt greenhouse gas protocol it is expected that their Scope 3 emissions calculations will be updated. It helps to have structured change logs to back these adjustments and methodology changes such that they can be demonstrated at the time of final assessment. Ensuring change logs and overall control discipline helps to ensure that the ESG disclosures are as accurate as financial report adjustments. This not only makes the process more transparent and accountable but also backs it up with appropriate audit trails that are assurance ready.

Versioning: How it helps

It is also important to incorporate the principle of versioning whenever there is any methodological change. By this we mean any changes to dimensions such as emission factors improvements in supplier data or even revisions in regulations which need to be reflected through change logs and appropriate versioning mechanism. This versioning mechanism should be able to explain why the historical figures have changed over time whilst preserving the historical calculation logic.

For instance, organizations who are part of science-based targets initiative are often found to be recalculating emissions baselines whenever acquisitions or divestitures of underlying

entities take place. As such these recalculations can tangibly impact historical emissions figures. Having an appropriate version control mechanism ensures that analysis is able to reconcile new numbers with historically reported figures and prospective investors are able to meaningfully interpret revisions to the new disclosures.

Reconciliation issues: where ESG data often breaks

Unlike financial reports ESG data is often a multi-disciplinary and a multi system artifact. By this we mean that ESG data are typically expected to originate from operations procurement systems HR databases, supplier disclosure and third-party data sets. As such it is important to establish reconciliation among the varied data sources while consolidating for the final disclosure. Some of the key reconciliation challenges are:

- **Boundary mismatches:** The data boundaries are typically varied. For instance, operational data may reside within the facility's boundary, while financial data is encapsulated within legal boundaries.
- **Time frame differences:** the frequency of capturing information also varies. For instance, operations data are found to be captured almost daily, while financial data is often at a more staggered interval, such as a monthly or quarterly basis.
- **Variability of information:** Scope 3 emissions data is dependent on supplier disclosure; on the other hand, the equality of supplier reporting varies greatly. As such, there is significant uncertainty in how the scope 3 estimates are calculated.

A typical example of reconciliation issues has been noted by major multinational organizations such as Unilever and Nestle, who have publicly acknowledged that most of their emissions occur in their supply chains. These emissions require aggregating data across thousands of suppliers who are spread over multiple regions and industries. As such, reconciling this varied data set with the internal procurement data creates significant complexity and data management challenges. These businesses, therefore, have invested significantly in engaging their suppliers and onboarding them to their own digital platforms in order to drive information traceability and verifiability.

Audit implications: Moving from disclosure to assurance

Auditors typically do not test the ESG metrics by simply reading sustainability reports. Instead, they look to verify the data source transformation logic that has been applied, controls and balances enforced, and governance of methodologies. The metric is expected to follow an end-to-end evidence chain - this comprises source data information, transformation logs, change logs with their versions, evidence of approvals, and reconciliation workings.

Assurance activities for ESG reporting are increasingly being found to follow standards such as the ISAE 3000, which stipulates that assessors seek sufficient and appropriate underlying evidence to support the management narrative. Lineage of the underlying data drives strong evidence of this.

Board-level implications and why investors care

Data lineage is at the heart of a governance issue. Managements are directly responsible for ensuring that the reported ESG metrics are reliable and defensible. This requires management to institutionalise ESG data governance frameworks, internal checks and balances, and integration across various internal systems, including those of financial reporting. The CFO and audit committee are expected to play critical roles in aligning the final disclosure with internal financial control disciplines. This often includes appropriate ESG data governance practices, reconciliation between financial and sustainability metrics, and oversight of ESG readiness.

In the current times, investors are found to be increasingly relying on ESG information as it captures important underlying risks, regulatory exposure, and operational resilience. Research from various academic bodies and investor guidelines suggests that the ESG data variability is largely caused by differences in underlying methodology and electricity stop in other words investors in today's world not only want to know what the companies report but how those numbers are produced, which stops have such ensuring traceability across the data improves investor confidence and reduces risk of being misrepresented or green-washed.

How to design a traceable ESG data architecture?

There are some best practices that an organization can use to deliver a traceable data architecture:

- Data lineage mapping: It is often helpful to have a visual map of how the ESG information is flowing across various systems and boundaries
- Automated change logs: systems should be designed such that change logs are automatically collected with the user ID and time stamp information
- Version control methodologies: It is important that data repositories preserve previous calculation logic with appropriate version numbers.
- Reconsideration frameworks: processes must be designed to ensure that the sustainability information can easily be reconciled with operational financial data
- Centralised storage of information: management must invest in a robust and resilient centralised storage of documentation related to ESG disclosure

ESG reporting is fast moving from narrative-based disclosures to regulated information that is reconcilable, traceable and verifiable. In this new environment the underlying credibility of information rests singularly on how traceable the underlying data is. As such data lineage becomes non-negotiable for organizations who want to explain how the numbers were produced and drive an end to end visibility; which is of utmost importance at the time of assessment to ensure outcomes are in alignment with management aspirations.