European Journal of Sustainable Development (2026), **15**, 4, 871-889 ISSN: 2239-5938
Doi: 10.14207/ejsd.2026.v15n4p871

From Traceability to Carbon Accounting: A Bottom-Up ELT Architecture for Product-Level Emissions Tracking under CBAM

By Pavan Savitha Manchegowda[1], Cosimo Di-Cecca[2], Nico Hanenkamp[3]

ABSTRACT:
The EU's Carbon Border Adjustment Mechanism (CBAM) has transformed carbon pricing from a theoretical policy debate into a hard financial liability. Consequently, treating environmental data merely as a compliance output is no longer practical. While supply chains restructure dynamically, traditional top-down ESG reporting relies on static averages. To transform sustainability metrics into a true Business Intelligence (BI) layer, this study presents a bottom-up Extract, Load, Transform (ELT) data architecture along with real-time dashboards for visualization. By enriching machine-level process data with component identification, we calculate product and machine-level sustainability KPIs. Electricity consumption and CO_2 emissions are derived either directly from the logged voltage and current values where available or via proxy models combining process duration with machine rated power constants. CO_2 emissions are calculated subsequently using the grid emission factor from the local energy provider. Visualizing these precise, real-time KPIs alongside operational parameters empowers managers to instantly assess environmental impacts without complex manual calculations, while also enabling the capture and analysis of highly granular, non-aggregated data at the point of generation. To generalize these findings, we introduce a 4-level Sustainability Data Maturity Model (SDMM), guiding facilities from estimate-based reporting toward precise carbon accounting. Ultimately, this architecture provides a scalable foundation for manufacturers to transition from reactive compliance to proactive, process-integrated and data-based sustainability management.

Keywords: Bottom-up carbon accounting, ELT data architecture, Digital traceability, Sustainability Data Maturity Model (SDMM), Carbon Border Adjustment Mechanism (CBAM)

1. Introduction

The European Union's (EU) Carbon Border Adjustment Mechanism (CBAM) became effective at the beginning of 2026. This development is crucial for manufacturing industries importing their products and raw materials from the non-European countries to the EU (Regulation (EU) 2023/956 (CBAM), 2023). This has transformed the carbon footprint of the product from a sustainability metric to a direct financial variable. From 2026 onwards, the importers of the covered goods must declare the emissions and provide the CBAM certificates with the prices for these certificates linked to EU Emission Trading System (ETS). In this context, the product-level carbon footprint (PCF) is no longer just environmental reporting but has become a part of cost, sourcing and traceability in supply chain management (Ubacht et al., 2023).

[1] Chair of Resource and Energy-Efficient Production Machines (REP), FAU Erlangen-Nürnberg, Germany
[2] Chair of Resource and Energy-Efficient Production Machines (REP), FAU Erlangen-Nürnberg, Germany
[3] Chair of Resource and Energy-Efficient Production Machines (REP), FAU Erlangen-Nürnberg, Germany

Manufacturing firms rely on top-down approaches for calculating the sustainability relevant KPIs for reporting, using the factory-level estimates for the energy consumption, utility bills and generic industry averages. This method becomes obsolete for the low-volume production units, where product variants, process routes and machine settings change frequently (Gan et al., 2023). As a result, the product-level sustainability data remain disconnected from the operational and production data, creating a gap between the production activity and environmental impact. This gap becomes significant not only from a product specific carbon footprint perspective but also for firms that need to collect or provide data for auditable sustainability reports aligning with the Corporate Sustainability Reporting Directive (CSRD) in accordance with the European Sustainability Reporting Standards (ESRS) requirements (United Nations Environment Programme, 2026). Furthermore, initiatives such as the EU's Digital Product Passport (DPP) aim to advance transparency and accessibility to product-related sustainability data by mandating selected industries to disclose information about product origin, materials, environmental impact, and waste disposal (Regulation (EU) 2024/1781 (ESPR), 2024).

Granular production data for components and products is available in production systems, but it is not used for detailed PCF calculations at the part or product-level. This fragmentation results from data being distributed across machines, local logs, and other enterprise systems in inconsistent formats. In low-volume production environments like the prototyping operations, this fragmentation is intense due to high process variability, frequent engineering changes, difficult traceability across multiple production and test environments (Gan et al., 2023). The process data that is used for quality assessment can be used for sustainability assessments at product and process levels provided, it is integrated into a structured data architecture. Without such integration, manufacturing firms may know what was produced, but cannot quantify the energy consumed or the emitted carbon from the individual machines or the products. This leads to reliance on aggregated, unspecific data for process optimization and sustainability reporting.

This paper addresses that problem by presenting a bottom-up Extract, Load, Transform (ELT) architecture that integrates traceability and sustainability analysis within a single layer. The architecture was implemented in the prototyping unit of a major Tier-1 automotive supplier in Germany and connects machine-level data collection, standardized data processing and dashboard-based visualization to part-level identification via Data Matrix Codes. The same digital thread enables retrieval of production histories and process parameters while also supporting sustainability key performance indicators such as process efficiency, estimated electricity consumption and CO_2 emissions. By converting operational data into analysable sustainability information, the architecture moves environmental accounting closer to the point of data generation.

The study makes three contributions. First, it shows how a practical industrial ELT architecture can transform fragmented machine data into a traceable and auditable digital thread for prototyping operations. Second, it demonstrates how electricity use and CO_2 emissions can be derived at machine, product and component-level using either direct telemetry or proxy-based estimation methods. Third, it introduces a four-level Sustainability Data Maturity Model (SDMM) that describes the transition of environmental impact assessment of manufacturing operations, from an unmeasured state (Past), to a process and product-level carbon accounting through proxy-based and direct calculations.

Together, these contributions position sustainability data not as a retrospective compliance output, but as a business intelligence layer for operational decision-making. The implication is that digital traceability and environmental accounting should be treated as complementary capabilities, not separate tasks.

2. Background and Related Work

The enforcement of Carbon Border Adjustment Mechanism (CBAM) by the European Union (EU), hits the international manufacturing supply chains directly. As an important step towards the climate action, the CBAM incorporates the cost of carbon by levying a tariff on the imported emission-intensive raw materials and semi-finished products to match the carbon price paid under the EU Emissions Trading System (ETS) (Nyaga, 2025). Recent economic evaluations show that the global manufacturers exporting to the EU face direct consequences as the CBAM implementation is rolled out and is expanding to include the scope 2 and scope 3 emissions (Li et al., 2026). The boundary conditions for the product-level greenhouse gas (GHG) accounting as well as the framework and reporting principles for bottom-up calculation are outlined by the World Resources Institute (World Resources Institute, 2011). Crucially, the severity of this regulation relies heavily on the manufacturer's data tracking capabilities, as it operates on a two-tier sanction system. The first tier relies on the carbon duty based on verified, actual product-level embedded emissions. If the manufacturer lacks the infrastructure required for the Monitoring, Reporting, and Verification (MRV) of the credible emissions data, they are subjected to a tier-two sanction. This tier reverts to application of a default value-based carbon price calculated on the basis of emission intensity of the ten worst-performing exporting countries as a worst-case scenario, which in turn builds a participation pressure. In summary, the EU CBAM turns the absence of granular data into a financial penalty, making bottom-up traceability of carbon footprint a critical economic necessity for global manufacturers (Nyaga, 2025).

Despite the clear requirement for a verifiable carbon accounting on product-level to fulfil the regulatory and financial compliance, current manufacturing environments face significant challenges in terms of traceability which arises due to poor integration between the industrial machinery and the high-level data management systems. Industry 4.0 and digitalization of manufacturing operations rely on interconnection between these systems, to facilitate the removal of the information silos, creating better end-to-end traceability of products across the entire value chain (Ghobakhloo, 2020). However, achieving this level of digital integration requires significant organizational and technical effort. Manufacturing systems are naturally characterised by the diverse, unstructured processes which in turn complicate the identification and acquisition of unified data (Schreiber & Metternich, 2022). The failure to resolve these data silos challenges the bottom-up, end-to-end transparency required to track component-level sustainability metrics and process histories across manufacturing (Ghobakhloo, 2020; Schreiber & Metternich, 2022).

A bottom-up methodology for identifying the sustainability relevant KPIs to reduce the environmental impact was presented by Glišić et al. It is based on the identification and improvement of process performance KPIs related to infeed centreless grinding process. The study focussed on the process KPIs such as grinding time, grinding

© 2026 The Authors. Journal Compilation © 2026 European Center of Sustainable Development.

power consumption and spark out time showing that reducing these parameters, can contribute to lower climate impact (Glišić et al., 2024). This study focused on a single machining operation and it did not address the challenge of heterogeneous data coming from different machines and operations. Furthermore, the focus was process-specific and did not extend to linking the KPIs to individual products or components. Thiede proposed using state-based calculations and machine learning methods to estimate the machine energy consumptions profiles across punching, CNC machining and weaving machines to identify the Overall Equipment Effectiveness (OEE). Although multiple machines types were validated, integrated multi-machine data pipelines or the component-level traceability were not addressed (Thiede, 2023). Wurm et al., developed a methodology for accurate prediction of Product Carbon Footprint (PCF) calculation in machining processes. They evaluated three established standards namely the ISO 14067, the GHG protocol, PAS 2050 and validated the approach on a single milling machine in a laboratory setting. While the methodology provides detailed guidelines for input and output stream mapping, it relies on manual data collection for each process (Wurm et al., 2025). Neither of the above studies addressed the challenge of integrating heterogeneous data arising from multiple machines simultaneously, nor the mapping of derived sustainability KPIs to individual products or components across production sequences.

Table 1: Comparison of related approaches to derive sustainability KPIs and product-level carbon footprint in manufacturing

Study	Architecture	Traceability Level	Calculation Method	Data Granularity	Validation Setting
Glišić et al. (2024)	Bottom-up process monitoring framework	Process level	Bottom-up process-parameter-based KPI selection	Single machining operation	Laboratory case: centreless grinding
Thiede (2023)	State-based energy estimation framework	Machine level	State-based estimation with machine learning	Multiple machine types — individual	Laboratory case: punching, CNC machining, weaving
Wurm et al. (2025)	Standard-based PCF methodology	Process level	Direct measurement and standard formulas (ISO 14067, GHG Protocol)	Single machining use case	Laboratory milling case study
Schreiber & Metternich (2022)	Data value chain framework	Process history level	Process mining and data value chain analysis	Single production system	Industrial conceptual demonstration

None of the reviewed approaches answers the component level traceability in a multi-machine production environment in a real industrial environment using a persistent identifier.

A significant challenge observed in small and medium scale manufacturing companies is the heterogeneous nature of machine data. Unifying this data is further challenged by the architectural limitations of traditional data integration pipelines. Data warehouses typically rely on the Extract, Transform, Load (ETL) or "schema-on-write" methodology, which is rigid and relies on sequential steps to extract the data from the source, transform it within a staging area into a structured format and finally load it into a repository. This method has its limitations in the modern manufacturing environments where the datasets are characterised by heterogeneous, semi-structured, and unstructured formats (Kancharla, 2025).

To overcome the disadvantages of the traditional ETL architecture, data architectures are shifting towards the Extract, Load, Transform (ELT) paradigm. This method reverses the traditional integration sequence by prioritizing data ingestion efficiency. It loads the raw, unprocessed data directly into the target repositories, such as cloud-based data lakes before any transformations are applied. By leveraging the elastic computing capabilities of modern cloud platforms, ELT allows complex transformations to be executed dynamically via distributed processing frameworks or powerful in-database SQL engines, effectively eliminating historical resource constraints (Kancharla, 2025).

While the architectures like ELT and data virtualization solve the backend bottleneck associated with data integration, the final outcome can only be quantified based on the effectiveness of the presented data to the decision makers. This final critical phase of 'Data exploitation' is to serve as the information base for the strategic decision-making process based on the visualization of the analytical output. Conventionally the monitoring and the forecasting methods have relied on static data and manual reporting processes (Osho et al., 2024). To overcome the limitations of conventional methods and to effectively manage the modern supply chain and manufacturing, the integration of Business Intelligence (BI) platforms—such as Microsoft Power BI—has become essential (De Castro, 2022). By transforming the raw ELT outputs into simple, open and real-time interactive dashboards, BI platforms simplify the data visualizations across the organization, making the critical information easily visible and shareable (Nabil et al., 2023).

3. Proposed Architecture: Bottom-up ELT for Sustainability

Existing manufacturing infrastructures generate rich process data at machine-level, yet these data remain fragmented, lack a uniform structure, and are disconnected from component-level traceability. To bridge this gap and enable product-level traceability, we propose an end-to-end Extract Load Transform (ELT) pipeline. The ELT approach was selected over traditional ETL architecture because of the heterogeneous, semi-structured nature of the machine data sources. ELT implementation requires loading raw data before transformation as it provides better flexibility compared to Extract Transform Load (ETL) pipelines (Kancharla, 2025). The architecture consists of five sequential layers:

© 2026 The Authors. Journal Compilation © 2026 European Center of Sustainable Development.

data generation, data collection and transfer, cloud storage, data processing and transformation and visualization. The granularity of the data is preserved at each layer while the necessary enrichment to derive the sustainability relevant information is carried out at the final transformation layer.

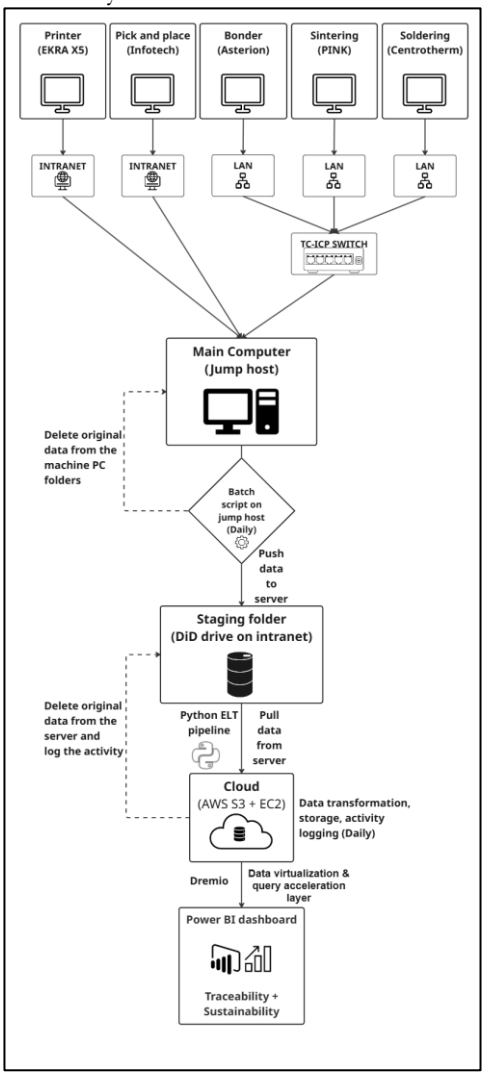

Figure 1: *Bottom-up ELT architecture for sustainability KPI derivation: five-layer data pipeline from machine-level generation to BI visualization*

3.1 Layer 1 — Data generation at machine-level

Each machine in the scope is capable of generating the process logs along with the detailed timestamps. Process data contain varied information like the actual values, set values for the temperature, pressure, gas flow rate, information regarding the semiconductor chips and their placement positions, process quality etc. This data is usually retrieved from the Programmable Logic Controller (PLC) of the machine and saved into

the machine computer used for input and visualization. Where the innate logging capabilities are absent, a software upgrade to include this feature without altering the operational UI of the machine was performed along with the machine supplier. The existing hardware of the machine was not altered in any way and where required, an additional piece of equipment like a hand-held scanner was installed to input the component identification data.

3.2 Layer 2 — Data collection and secure transfer

Raw data files from the local storage of machines are transferred to cloud via a dedicated on-premise computer (jump host), which acts like a secure interface between the Operational Technology (OT) and the Information Technology (IT) domain. The data from the machines are accessed from the jump host either through LAN connections for old machines that cannot be secured or through intranet connections for the new machines. Dedicated batch scripts on the jump host PC, running according to a schedule, copy the data and also take care of clearing the machine memory once the data is securely copied. The data from the jump host is transferred to cloud storage via a cloud-initiated pull based on the zero-trust architecture principle required in industrial cybersecurity frameworks.

3.3 Layer 3 — Cloud storage: Raw data lake

The raw data files from the jump host are stored unmodified into a cloud-based object storage layer functioning as Data Lake. The ELT principle is followed, where the "load first, transform later" is observed, to allow the full fidelity of original machine data. On data lake a two-tier directory structure is used to separate the raw data from the transformed data enabling clear data lineage. A unique component identifier information which comes in from the Data Matrix Code (DMC) engraved on each component for individual component tracking is present in the machine log and serves as primary key across all the datasets.

3.4 Layer 4 — Data processing and transformation

Machine-specific scripts scheduled to run as instances take in the raw data as input and perform the data normalization, timestamp parsing and conversion of waveform data into suitable time series. Each output file produces a standard output schema containing the mandatory columns such as DMC, timestamp and source machine alongside other process-specific data from the machine. A central traceability table is constructed at this stage by appending the standard output from all the machines. This acts as a factless fact table in a hybrid star schema, combining a centralised traceability table with machine-specific dimension tables. Sustainability relevant KPI calculations are applied at this stage as derived measures, converting raw data inputs into energy and emission estimates.

3.5 Layer 5 — Visualization and reporting

The processed dataset is imported into a business intelligence platform via a cloud SQL engine, to build real-time interactive dashboards. Power BI was selected as the visualization platform based on its native integration with the Microsoft enterprise stack already deployed at the implementation site. This ensured immediate accessibility without additional infrastructure. This decision aligns with the findings that BI platform adoption

© 2026 The Authors. Journal Compilation © 2026 European Center of Sustainable Development.

is most effective when embedded within the organization's existing digital ecosystem (De Castro, 2022). The visualization layer is structured into two views, machine-level dashboards providing the operational and sustainability KPIs that can be drilled down to a single process level and a master traceability report which enables component-level traceability via DMC look up. This dual view format ensures that the same data infrastructure enables process quality and environmental KPI visualizations.

Implementation of this pipeline is done through Python-based data transformation scripts running on AWS EC2 compute instances. Both raw and processed data are stored in AWS S3 object storage organised in a two-tier directory structure. Data virtualization and query acceleration are performed through Dremio, which connects the processed datasets to the Power BI visualization layer via a cloud SQL engine. For the security of the data systems, a zero-trust architecture principle is followed. Data transfer is initiated by a cloud pull through Python scripts rather than a push from the on-premise Jump Host. The machine networks are isolated through separate LAN and intranet connections for legacy and modern machines respectively. The data is governed through a versioned parameter table that explicitly documents all KPI calculations, constants, and emission factors, ensuring the auditability of every derived metric. The architecture is vendor-agnostic — any cloud object storage service can be utilised in place of AWS S3, Dremio can be substituted by any cloud-compatible SQL engine, and Power BI with any BI platform that supports live cloud data connections. The only implementation-specific elements are the machine-specific transformation scripts, which require modification based on the machine data output format. Any manufacturer seeking to replicate this pipeline requires a dedicated Jump Host PC connected to the machine network via LAN or intranet, a cloud service account from any provider, a Python runtime environment, and a BI platform license. No dedicated hardware instrumentation beyond what is commercially available is required.

4. KPI Calculation Logic

Sustainability KPIs are derived from the data transformation layer, before it reaches the visualization platform (Power BI). The KPIs are organised into two categories: directly measured and proxy-based indicators.

4.1 Directly measured indicators

The foundational information in aggregating the KPIs down to the product-level, is the process cycle duration for each process run the product undergoes in the production. For event-driven machines, the duration is calculated as the difference of the process start and end timestamps against the product DMC. For batch production machines, multiple components are processed together, the duration is calculated based on the maximum elapsed time value within the process run indicator. Some machines where the process quality information is available directly in the process or machine logs, process quality visualization is built directly to indicate if the run was good or bad. For machines without this binary information, pre-set values were used for the comparison between the measured and the set values. These quality indicator KPIs directly indicate the

sustainability of the production as the defect and rework indicate wastage of the material and energy.

4.2 Direct and proxy calculations

For machines where current and voltage values are available in the process logs, electricity consumption is calculated directly without relying on rated power constants. For the Asterion wire bonder, instantaneous power is derived as the product of the logged voltage (V) and current converted from milliamps to amps (mA ÷ 1000). Energy consumption per process run is then calculated by integrating this power value over the measured process duration. For machines where the direct data regarding the energy consumption or the electricity consumption is not available, it is calculated as the product of the average rated power of the machine and the measured usage duration.

$$\textit{Total Electricity (kWh)} = \textit{Usage Duration (hours)} \times \textit{Average Rated Power (kW)} \quad \text{(Equation 1)}$$

where Usage duration is the measured process cycle time derived from the machine timestamp of a process. Average Rated Power is the machine-specific constant in kilowatts which is taken from the manufacturer specifications.

CO_2 emissions are subsequently derived by applying the grid emission factor to the estimated electricity consumption. The CO_2 factor information is taken directly from the local electricity provider.

$$\textit{CO}_2\textit{ Emissions (kgCO}_2\textit{eq)} = \textit{Total Electricity (kWh)} \times \textit{CO}_2\textit{ Emission Factor (kgCO}_2\textit{/kWh)}$$

(Equation 2)

where the CO_2 Emission Factor is expressed in $kgCO_2$ per kWh and this value is given by the regional electricity provider as an annual energy mix disclosure.

The proxy-based estimation introduces uncertainty at two levels. First, the variation between the machine's rated power and the actual power draw under operating conditions. Second, the usage of annual average grid emission factor to calculate the CO_2 emissions, as the actual grid emission varies by the input to the power plant at the time of power draw (based on whether the power production happened through renewable or non-renewable source and the type of fuel if it is non-renewable). Although a dedicated uncertainty analysis falls outside the scope of the present work, drawing on similar proxy-based energy studies allows for a reasonable approximation of the expected error margins. In industrial environments, the discrepancy between rated and actual machine power introduces estimation uncertainty, the magnitude of which depends on the specific machine type, load profile, and operating conditions. Furthermore, relying on an annual average CO_2 emission factor introduces an additional source of variance arising from temporal fluctuations in grid emission intensity that are not captured by annual average factors. Together, these two variables represent the primary sources of estimation uncertainty underlying the calculated proxy-based KPIs. Direct voltage and current output

© 2026 The Authors. Journal Compilation © 2026 European Center of Sustainable Development.

in the data as in the case of the Asterion wire bonder machine completely eliminates the uncertainty arising from Average Rated Power usage, producing the total electricity consumption values with instrumental accuracy.

A sensitivity analysis of the component-level carbon footprint reveals that the process duration calculation has a direct influence on the proxy calculation, as it directly scales the electricity estimate linearly. Since this calculation is made from the direct process data, the probability of error is less. The rated power constant is the second most influential variable. As shown in Equation 1, it directly influences the total electricity consumption value. For example, a 20% overestimate in rated power produces a 20% overestimate in the electricity consumption. The grid emission factor influences the CO_2 emission calculation proportionally as shown in Equation 2. Batch allocation, as applied to the EKRA printing process, introduces additional sensitivity because the per component footprint is inversely proportional to batch size. Consequently, a smaller batch size results in a higher per-component emissions. As shown in Table 1, the EKRA printing process contributes only 0.014 kg CO_2eq per component from a batch of 216 parts, demonstrating that batch allocation significantly reduces the per-unit environmental impact of shared process steps.

This proxy-based approach to derive the product-level carbon footprint is based on the conditions and the calculation principles as laid out in the ISO 14067:2018 for quantifying and communicating the product carbon footprints (International Organization for Standardization, 2018). Using a supplier-specific emission factor rather than national averages improves the accuracy and local relevance of the CO_2 calculation. These constants and calculation parameters are stored in a dedicated parameter table within the transformation layer, ensuring that the assumptions underlying each KPI calculation are versioned, transparent, and auditable under CSRD requirements.

5. Implementation Case Study: Automotive Supplier Prototyping Unit

5.1 Industrial context

The architecture was implemented and validated at the prototyping unit of a major Tier-1 automotive supplier in Germany, specialising in drivetrain and electrification components. The unit produces power electronics modules for electric vehicle applications through processes including screen printing, pick-and-place component placement, sintering, vacuum soldering, and wire bonding. The unit represents a high-mix, low-volume production environment with frequent process variation, multiple machine vendors, and no pre-existing unified data infrastructure, making it a suitable case for the challenges outlined in section 2.

5.2 Machine digitalization

Five of the six identified machines were successfully incorporated into the architecture, namely the EKRA X5 screen printer, Infotech pick-and-place machine, PINK SIN 200+ sintering press, Centrotherm vacuum soldering system, and Asterion wire bonder. Two of these required the procurement of additional software for the data logging along with DMC information of the components. Five machines logged the process data in four distinct formats, structured CSV, semi-structured SPC text files,

proprietary binary format, and high-frequency waveform CSV exports. This heterogeneous nature motivated the ELT approach.

5.3 Data pipeline performance

The data collection architecture using the on-premise jump host demonstrated a stable daily operation across all the five machine sources. The cloud-initiated data pull from the jump host to the data lake achieved a 100% successful data ingestion rate across the observed period. The automated clean-up of the machine memory after the ingestion manages the machine storage constraints. The transformation scripts were successful in parsing all the four file types into a standard CSV format. The most complex transformation was applied to Centrotherm binary parser and the Asterion waveform reshaper, both of which were converted from their proprietary format into a flat schema CSV format.

5.4 Traceability results

The master traceability report page was able to demonstrate the end-to-end traceability of each component. As an example, querying the component with DMC REP202530073, gave a chronologically ordered list of the timestamps and machines on which this component was processed, spanning multiple processes. A drill through navigation enabled machine-level process and sustainability KPI visualization for each event. The conditional formatting of the navigation buttons provides visual indication of the machines on which the part was processed without having to manually cross reference.

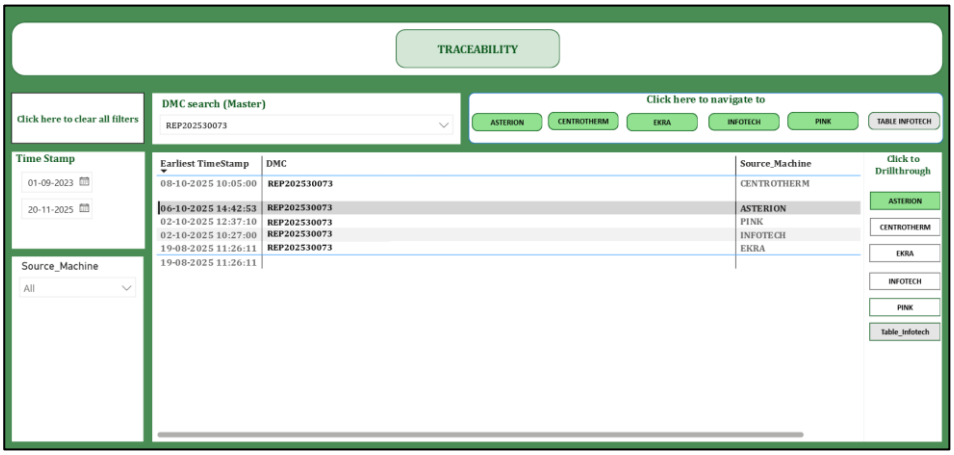

Figure 2: *The screenshot of the master traceability report filtered to a single DMC (REP202530073); the interface shows the component's complete production history across multiple machines. A direct access to the machine-level sustainability and process KPIs is enabled through drill through buttons for each process event.*

5.5 Sustainability KPI results

First Pass Yield for the PINK sintering press was measured at 78%. The Infotech machine recorded an out-of-specification deviation rate of 21.87%, identifying it as a

© 2026 The Authors. Journal Compilation © 2026 European Center of Sustainable Development.

primary source of material and energy waste. Proxy-based electricity and CO_2 estimates were calculated using the N-ERGIE[4] 2024 grid emission factor of 354 gCO_2/kWh.

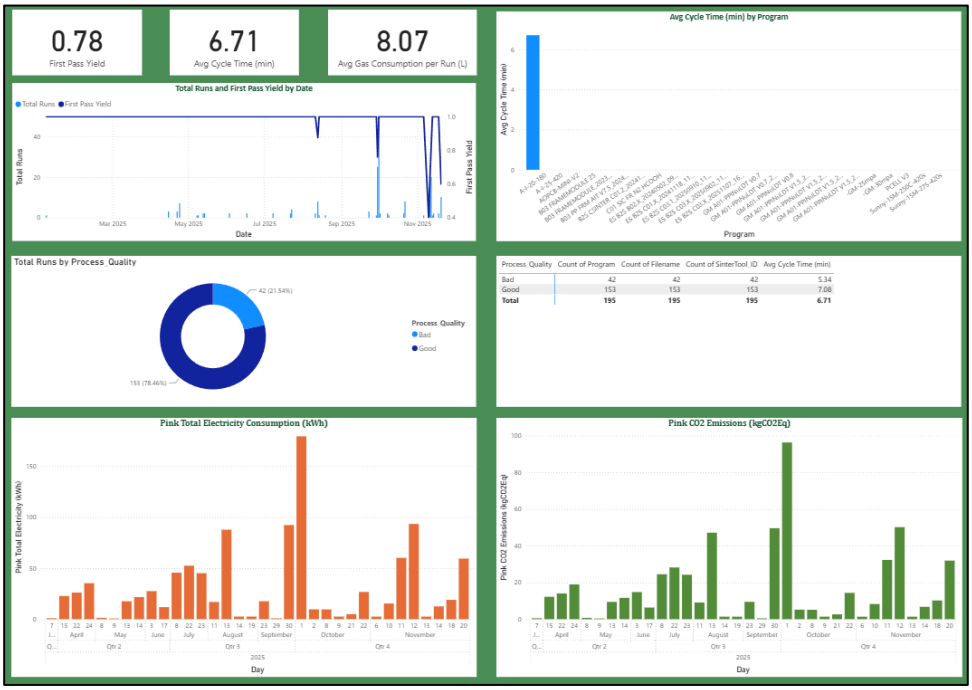

Figure 3: *Power BI dashboard visualizing real-time process quality and CO_2 emissions of a single machine (PINK). The top panels show the traditional manufacturing KPIs, including First Pass Yield and Cycle Time, alongside process quality distributions. The bottom visualizations show the electricity consumption and associated CO_2 emissions drilled down to a daily granularity.*

A component-level carbon footprint case study was conducted for substrate DMC REP202530073 across five sequential processes. This component underwent five sequential manufacturing stages, printing (EKRA), pick-and-place (Infotech), sintering (PINK), wire bonding (Asterion) and vacuum soldering (Centrotherm). The architecture successfully aggregated energy and CO_2 contributions from each stage into a single component-level carbon footprint. This demonstrates that deriving a product-level carbon footprint from operational data without dedicated energy metering is achievable and provides a pathway towards the CBAM compliant verified, product-level embedded emissions data.

[4] Using local emission factors of N-ERGIE Aktiengesellschaft (~2,500 employees, major electricity/gas supplier in Middle Franconia) as the actual energy supplier of the case company.

Table 2: Carbon footprint aggregation for the component with DMC REP202530073 across five sequential manufacturing processes

Process	Machine	Estimated CO_2 emissions ($kgCO_2eq$)	Note
Printing	EKRA	0.014	Calculated from a batch of 216 parts (Total emission - 3.09 $kgCO_2eq$)
Pick and place	Infotech	1.36	-
Sintering	Pink	1.37	-
Bonding	Asterion	0.05	Calculated from a batch of 2 parts (Total emission - 0.10 $kgCO_2eq$) (Calculated from direct V×I measurement)
Vacuum Soldering	Centrotherm	2.88	-
Total Footprint		**5.67 $kgCO_2eq$**	

Note: CO_2 = carbon dioxide; $kgCO_2eq$ = kilograms of carbon dioxide equivalent; EKRA, Infotech, PINK, Centrotherm, and Asterion are third-party machine manufacturer names. Printing emissions are batch-allocated: total batch emissions divided equally across 216 components. Bonding emissions are batch-allocated across 2 components and calculated using direct voltage × current (V×I) measurement rather than proxy estimation. A dash (–) indicates no additional methodological note applicable to that process.

5.6 Summary of implementation outcomes

The industrial case study confirms that the proposed architecture can be deployed in production environments with heterogeneous and legacy machine infrastructures. Unlike the implementations of Glišić et al. (2024) and Thiede (2023) where sustainability KPIs were derived at a single machine or single process level. While, Wurm et al. (2025) validated the PCF calculation for a single machine operation in a laboratory setting, the presented architecture in our study integrates five heterogeneous machines in a real industrial environment, linking all derived metrics to individual components through a persistent DMC identifier — a level of granularity not demonstrated in any of the reviewed studies. In terms of scalability, the ELT approach with cloud-based storage is capable of accommodating additional machines through machine-specific transformation scripts without modifying the underlying pipeline. Whereas the standard-based methodologies of

© 2026 The Authors. Journal Compilation © 2026 European Center of Sustainable Development.

Wurm et al. (2025) and the monitoring frameworks of Glišić et al. (2024) do not address multi-machine integration. In terms of deployment, all reviewed studies validate their approaches in laboratory or conceptual settings, while this work demonstrates implementation in an operational industrial environment with legacy machine infrastructure, heterogeneous data formats, and real production constraints. The implementation also demonstrates that a single data pipeline system using the machine process data can be used to derive the process and sustainability relevant information, without requiring separate systems.

6. The Sustainability Data Maturity Model (SDMM)

The observed findings and the limitations from our proposed architectural implementation, motivate us for the formulation of a four-level Sustainability Data Maturity Model (SDMM). This model describes the progression of manufacturing industries to measure, attribute and report the environmental impact of their production operations. It serves as a roadmap, enabling organizations to locate their current state and to plan their investments towards the primary data-based sustainability accounting.

Level 1 — Unmeasured (The past)

At Level 1, the environmental impact of the manufacturing operations remains unknown at a process level. Quantitative sustainability data is typically based on the high-level factory utility estimates, monthly electricity invoices, gas bills and other aggregated operational metrics. There is no possibility to attribute the consumption of resources to machines, process and product-level. This is the standpoint for most of the small and medium scale manufacturers. While large manufacturers disclose corporate sustainability at Level 1, they are still structurally incapable of providing the product-level carbon footprint relevant to CBAM compliance. This architectural gap is what the current research directly addresses.

Level 2 — Process data and proxy-based (The present)

At Level 2, the machine-level process data is systematically collected and aggregated to derive the sustainability relevant KPIs. Rather than relying on the factory-level aggregates, the energy consumption is estimated at process level on machines. CO_2 emissions are derived by applying localised grid emission factors. This level can be achieved by the digitalization of existing machine data without the requirement for a dedicated energy metering hardware as demonstrated in the industrial implementation presented in this article. Critically, when the process data is linked to a persistent component identifier like the Data Matrix Code (DMC), this level directly enables product-level carbon footprint aggregation across multi-stage production processes. The proposed architecture in this paper offers a practical and immediately deployable solution for manufacturers operating in low instrumentation environments. This level also serves as the data foundation for the higher data maturity model.

Level 3 — Direct measurement (The near-future)

This level replaces the proxy-based estimation with the direct real-time measurements for the energy consumption of the machine. This can be achieved with the installation of dedicated power and gas sensors at machine-level. This enables the logging of energy consumption continuously instead of inferred and rated constants, eliminating the possibility of estimation uncertainty. This level provides for precise, verifiable bottom-up production-related data for official reporting, including CBAM and third-party audits. The transition from Level 2 to Level 3 requires investment in terms of hardware, like smart meters and integration of sensor data streams into existing ELT pipeline. The rest of the data architecture for the collection, storage, transformation and visualization can be kept unchanged, which means that the organizations that have implemented Level 2 are prepared for this transition. Notably, machines that readily log the voltage and current data (like Asterion wire bonder), enable direct energy consumption calculation within level 2 architecture, partially closing the gap to level 3 without additional hardware.

Level 4 — Predictive and optimised (The far-future)

At Level 4, the sustainability data information transitions from descriptive and reporting to predictive. Machine learning and AI models can be incorporated using the Level 2 and Level 3 data sets, to forecast energy consumption, predict process deviations before they generate flawed products/scrap. This level advances from the reporting domain to a proactive, environmentally accountable level. Our architecture enables progressing to this level through the usage of cloud-based storage and the distinctive separation of raw and processed data on cloud. A step towards forward compliance. The SDMM is summarised in figure 4.

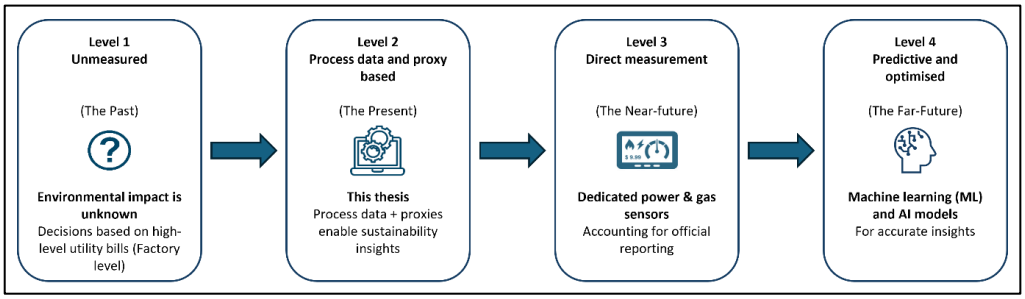

Figure 4. *The Sustainability Data Maturity Model (SDMM): Progression from unmeasured facility-level estimation (Level 1) to predictive AI-driven optimisation (Level 4). The architecture presented in this paper operationalises Level 2.*

7. Discussion & Implications

From compliance cost to business intelligence

The proposed architecture repositions sustainability reporting from a compliance requirement to a business intelligence layer. The pipeline developed to enhance the component and process traceability also helps in aggregating the sustainability KPIs. In this context, the marginal cost of adding the environmental reporting layer to the existing

© 2026 The Authors. Journal Compilation © 2026 European Center of Sustainable Development.

digitalization investment is minimal. This is achieved either by layering already existent process data or by implementing cost-efficient sensors and instrumentation. Organizations can treat traceability and sustainability as an integrated and unified data problem instead of building separate systems.

The CBAM financial argument for data investment

Organisations advancing across the SDMM stages, directly reduce their financial exposure under CBAM. Organisations at Level 1 face CBAM's default and higher carbon price calculated based on the emission intensity of the 10 worst performing countries. The difference in cost between this penalty and the verified product-level emissions represent the return on investment for advancing to Level 3 and Level 4 of the SDMM maturity. Investment in manufacturing data infrastructure is not merely an expenditure, but an operational excellence tool for controlling cost-efficiency, supply chain risk mitigation strategy and sustainability controlling (World Economic Forum, 2022).

Scalability and transferability

The architecture was validated in a prototyping unit consisting of five machines with four distinct data formats. The environment posed significant challenges in terms of process variability and data heterogeneity. The successful integration of these varied sources into a single pipeline demonstrated the architecture's scalable potential. By contrast, in a large-scale production or series manufacturing environment, process routes are more consistent and data volumes are higher. The same architecture can be scaled by adding machine specific scripts without modifying the pipeline structure. Only additional cloud resource needs to be allocated to accommodate the increased data volumes.

Limitations

The proxy-based estimation logic for power consumption on machines where direct measurement data is not available poses a risk of estimation uncertainty with reference to rated and actual power draw, as detailed in Section 4.2. The averaging of the power constant for this proxy calculation depends on the availability and volume of historical data. Furthermore, such KPIs indicate post-process analysis only and cannot support real-time reactive quality interventions where immediate corrective action is required at the point of deviation. The CO_2 emission factor supplied by the electricity provider is a static yearly average rather than a real-time grid value, introducing an additional variance. A key architectural limitation is the reliance on a single jump host, representing a possible point of failure in the data collection layer. This is a trade-off between the security, cost and redundancy that must be addressed in a full-scale production environment. The exclusion of the SMT oven, following a cost-benefit assessment that determined the software module cost of approximately €10,000 to be disproportionate relative to its data output value, created a gap in the digital thread and limits the completeness of the component-level carbon footprint that must be addressed in the future implementations. To address the limitations and to improve the reliability and generalisability of the present findings in full-scale production environments, increasing redundancy at the jump host level by an additional PC to fall back in case of failure can be incorporated. The validation presented in this study is limited to a single

prototyping site with a high-mix, low-volume production environment. While this represents a challenging implementation context that supports the argument for scalability, the lack of replication across multiple production sites and different machine configurations limits the generalisability of the findings. To strengthen the arguments considerably, longer observation periods and validation in series manufacturing environments are recommended as priorities for future work.

8. Conclusion

This paper presented a bottom-up ELT architecture to derive the product-level sustainability data. The sustainability KPIs were derived directly from the process data generated by the individual machines, which is linked to individual components through persistent identifiers. The architecture was validated at the automotive supplier's prototyping unit through successful integration of five machines with heterogeneous data sources, enabling process and component traceability alongside the sustainability indicators. The key finding from this implementation is that, product-level carbon footprinting is achievable using the existing machine-level process data when collected, standardized, and linked through unique component identifiers. The proxy-based logic demonstrated that the limitations associated with the data can be solved by incorporating the averages for rated machine power and the CO_2 factor from the local energy provider. Process duration combined with rated machine power and a local CO_2 emission factor proved sufficient to derive energy and emissions estimates attributable to individual components. The SDMM serves as a roadmap for manufacturing organizations to assess where they stand and to advance from facility level estimates to verified product-level carbon accounting. The absence of this granular emission data is a quantified financial liability arising from CBAM and no longer just a reporting gap. Future work should focus on four key aspects. First, the integration of real-time smart metering infrastructure at machine-level would eliminate the uncertainty associated with proxy-based calculations and advance the implementation to SDMM Level 3. This will enable the gathering of sustainability data at sufficient precision for formal CBAM declarations and also for third-party audits. Validation of the proxy values against the independent measurement devices would provide quantified accuracy for the proxy-based KPI values reported in this study. Second, replicating the data architecture presented in this study in a series manufacturing environment would validate the scalability beyond prototyping context. This would also involve testing the data pipeline's performance under higher data volumes and a larger machine count. Third, an economic assessment of the implementation cost, maintenance effort, and the return on investment through CBAM penalty avoidance would strengthen the business case for the adoption among manufacturing SMEs operating without dedicated sustainability infrastructure. Fourth, the formalization of the bottom-up KPI derivation methodology across different industrial sectors, validated across varied manufacturing processes and industries, would help extend the applicability of the SDMM beyond the automotive electronics context demonstrated in this study.

© 2026 The Authors. Journal Compilation © 2026 European Center of Sustainable Development.

References

De Castro, J. P. A. (2022). *Digital transformation BI tools integration with CRM and sales data* [Master's thesis dissertation]. Universidade de Aveiro, Portugal. https://www.proquest.com/openview/7f9359f16926fec3c4c785ee61e76ad5/1?cbl=2026366&diss=y&pq-origsite=gscholar

Gan, Z. L., Musa, S. N., & Yap, H. J. (2023). A review of the high-mix, low-volume manufacturing industry. *Applied Sciences*, *13*(3), 1687. https://doi.org/10.3390/app13031687

Ghobakhloo, M. (2020). Industry 4.0, digitization, and opportunities for sustainability. *Journal of Cleaner Production*, *252*, Article 119869. https://doi.org/10.1016/j.jclepro.2019.119869

Glišić, M., Veluri, B., & Ramanujan, D. (2024). A bottom-up methodology for identifying key performance indicators for sustainability Monitoring of Unit Manufacturing Processes. *Sustainability*, *16*(2), 806. https://doi.org/10.3390/su16020806

International Organization for Standardization. (2018). *ISO 14067:2018(en), Greenhouse gases — Carbon footprint of products — Requirements and guidelines for quantification.* ISO. https://www.iso.org/obp/ui/en/#iso:std:iso:14067:ed-1:v1:en

Kancharla, S. S. R. (2025). ETL vs ELT: Evolving approaches to data integration. *Sarcouncil Journal of Engineering and Computer Sciences*, *4*(7), 1065–1074. https://doi.org/10.5281/ZENODO.16405767

Li, M., Sun, Y., Xia, Y., Yang, Z., Chen, C., Yuan, Y., & Wang, P. (2026). The impacts of Carbon Border Adjustment Mechanism (CBAM) on international trade and policy responses: From an economic and environmental equity perspective. *Energy Policy*, *210*, 115014. https://doi.org/10.1016/j.enpol.2025.115014

Nabil, D. H., Rahman, M. H., Chowdhury, A. H., & Menezes, B. C. (2023). Managing supply chain performance using a real time Microsoft Power BI dashboard by action design research (ADR) method. *Cogent Engineering*, *10*(2), Article 2257924. https://doi.org/10.1080/23311916.2023.2257924

Nyaga, B. K. (2025). Turning data absence into a sanction: how the EU CBAM operates as a climate club to enforce carbon pricing participation. *The Journal of World Energy Law & Business*, *18*(4), Article jwaf026. https://doi.org/10.1093/jwelb/jwaf026

Osho, G., Omisola, J., & Shiyanbola, J. (2024). An Integrated AI-Power BI Model for Real-Time Supply Chain Visibility and Forecasting: A Data-Intelligence Approach to Operational Excellence. *International Journal of Advanced Multidisciplinary Research and Studies*, *4*, 1463–1480. https://doi.org/10.62225/2583049X.2024.4.6.4047

Regulation (EU) 2023/956 of the European Parliament and of the Council of 10 May 2023 establishing a carbon border adjustment mechanism, Official Journal of the European Union (2023). https://eur-lex.europa.eu/eli/reg/2023/956/oj

Regulation (EU) 2024/1781 of the European Parliament and of the Council of 13 June 2024 establishing a framework for the setting of ecodesign requirements for sustainable products, amending Directive (EU) 2020/1828 and Regulation (EU) 2023/1542 and repealing Directive 2009/125/EC, 2024/1781 Official Journal of the European Union (OJ L) (2024). https://eur-lex.europa.eu/eli/reg/2024/1781/oj/enga

Schreiber, M., & Metternich, J. (2022). Data value chains in manufacturing: Data-based process transparency through traceability and process mining. *Procedia CIRP*, *107*, 629–634. https://doi.org/10.1016/j.procir.2022.05.037

Thiede, S. (2023). Advanced energy data analytics to predict machine overall equipment effectiveness (OEE): A synergetic approach to foster sustainable manufacturing. *Procedia CIRP*, *116*, 438–443. https://doi.org/10.1016/j.procir.2023.02.074

Ubacht, J., Schmid, J., Rietveld, E., Lennartz, J., Rukanova, B., & Tan, Y.-H. (2023). Data sharing arrangements for monitoring in the EU circular economy: The case of CBAM and steel import for the EU automotive sector. In *EGOV-CeDeM-ePart conference.* CEUR-WS.

United Nations Environment Programme. (2026, May 4). *European Sustainability Reporting Standards (ESRS) – United Nations Environment – Finance Initiative*. https://www.unepfi.org/impact/interoperability/european-sustainability-reporting-standards-esrs/

World Economic Forum. (2022). *The data-driven journey towards manufacturing excellence*. https://www3.weforum.org/docs/WEF_The_Data Driven_Journey_Towards_Manufacturing_Excellence_2022.pdf

World Resources Institute (2011). Product life cycle accounting and reporting standard. https://ghgprotocol.org/sites/default/files/standards/Product-Life-Cycle-Accounting-Reporting-Standard_041613.pdf

Wurm, S., Tschepe, T., Petrovic, O., & Herfs, W. (2025). Developing a methodology for accurate product carbon footprint calculation in machining processes. *Procedia CIRP*, *135*, 1308–1313. https://doi.org/10.1016/j.procir.2024.12.148

© 2026 The Authors. Journal Compilation © 2026 European Center of Sustainable Development.