Vol. 1, No. 1, pp.: 33-44, 2023

CarbonLedgerDB: A Structured Database for Supply-Chain Carbon Accounting and Green AI Analytics

Jingjing Luo[1], Haoran Xu[2], Wei Fang[3,*]

[1] School of Management Science and Engineering, Nanjing University of Finance and Economics, Nanjing 210023, China

[2] Department of Information Systems and Decision Sciences, Shandong University of Finance and Economics, Jinan 250014, China

[3] School of Business Administration, Guizhou University of Finance and Economics, Guiyang 550025, China

** wei.fang@gzufe.edu.cn*

Article Information	
Received	18 October 2022
Accepted	29 February 2023
DOI	https://doi.org/10.63646/datamind.2023.010104

Abstract

Carbon accounting across global supply chains demands unprecedented data infrastructure: procurement records, logistics events, emission factors, and product-batch genealogies must be integrated and auditable at scale. Existing public databases address isolated aspects of this challenge but lack an integrated, versioned, and AI-ready architecture for end-to-end carbon lifecycle accounting. This paper introduces CarbonLedgerDB, a relational database system designed to support supply-chain carbon accounting and Green AI analytics. CarbonLedgerDB organizes five core entity types—Supplier, ProductBatch, EmissionFactor, LogisticsEvent, and CarbonRecord—into a formally specified schema with foreign-key traceability across Scope 1, 2, and 3 emission categories. The database ingests data from multiple heterogeneous sources including enterprise resource planning feeds, IoT sensor streams, logistics APIs, and curated emission-factor libraries. A structured quality-control layer manages missing values, duplicates, unit harmonization, and audit versioning. Experiments on a sample of 446 supplier entities, 12,800 product batches, and 31,200 carbon records demonstrate a mean carbon re-computation error of 2.3%, emission-factor coverage of 91.4%, three-tier supplier traceability of 87.6%, and audit consistency of 96.2%. CarbonLedgerDB is made openly available with a documented API and reproducible experiment notebooks, providing a reusable infrastructure for regulatory compliance analytics, machine learning model training, and organizational carbon intelligence.

Keywords: ***Carbon accounting; supply chain; emission factor; relational database; green ai; lifecycle assessment; sustainability analytics; open data***

1. Introduction

Supply chains are responsible for the majority of corporate greenhouse gas emissions. Studies repeatedly confirm that Scope 3 upstream and downstream activities account for more than 70 percent of the total carbon footprint of manufacturing and retail enterprises, yet these emissions remain among the most poorly measured and reported in corporate disclosure practice (Huang et al., 2009; Sarkis et al., 2011). The challenge is fundamentally one of data infrastructure: an organization seeking to compute its full lifecycle carbon inventory must integrate procurement contracts, logistics records, production schedules, energy consumption logs, and third-party emission factors across multiple tiers of often poorly documented suppliers (Daubé and Noury, 2022; Rao and Holt, 2005). Without an explicitly designed database architecture, each calculation remains ad hoc, non-reproducible, and impossible to audit at the required regulatory standard.

The regulatory context intensifies these demands. The European Corporate Sustainability Reporting Directive (CSRD), the U.S. Securities and Exchange Commission climate disclosure rule, and the ISO 14064 standard collectively require that organizations report auditable, verifiable, and methodologically consistent carbon inventories. The ISO 14040 and 14044 standards for life cycle assessment (LCA) further specify that emission calculations must be traceable to primary or secondary data sources with documented uncertainty (ISO, 2006; Rebitzer et al., 2004). Meeting these requirements without a purpose-built data infrastructure is increasingly untenable as supply chain complexity grows and regulatory scrutiny intensifies.

A second pressure comes from the Green AI research agenda. Training, fine-tuning, and deploying large machine learning models for supply chain optimization, demand forecasting, route planning, and sustainability reporting generates substantial computational carbon costs (Strubell et al., 2019; Schwartz et al., 2020). Benchmarking and reducing the carbon footprint of AI systems requires the same underlying data infrastructure that carbon accounting demands: structured emission factors, compute-energy conversion records, and auditable calculation logs. CarbonLedgerDB is designed to serve both corporate sustainability accounting and AI-carbon benchmarking workflows through a unified schema.

This paper makes three contributions. First, it presents a formally specified relational database schema for supply-chain carbon accounting, covering master data (suppliers, products, emission factors), transactional data (logistics events, production batches), and derived records (carbon computations, audit logs). Second, it describes the data pipeline connecting heterogeneous raw sources to a quality-controlled, versioned database. Third, it reports quantitative validation experiments assessing re-computation error, factor coverage, traceability completeness, and audit consistency on a multi-sector sample. The database is made openly available with documented interfaces, enabling reproducible research and regulatory-grade carbon intelligence.

2. Database Gap and Use Cases

Several existing databases partially address supply-chain carbon accounting, but each has structural limitations that prevent direct use as an end-to-end carbon ledger. The ecoinvent database provides industry-leading background life cycle inventory data with over 18,000 activity datasets, but its commercial licensing model and proprietary exchange format restrict integration into reproducible open pipelines (Weidema et al., 2013). The Global Emission Factor Database (GEFD) maintained by the International Energy Agency offers sectoral and country-level emission intensities but lacks product-batch or supplier-specific traceability fields. Environmentally Extended Input-Output (EEIO) databases such as EXIOBASE 3 (Stadler et al., 2018) and USEEIO provide macro-level supply-chain footprints but cannot resolve individual procurement transactions or logistics legs at the product and batch level.

At the enterprise level, ERP systems such as SAP S/4HANA and Oracle Fusion Cloud maintain procurement and logistics transaction records but are not designed to store emission factors or to compute and version carbon records automatically. Carbon management software platforms such as Salesforce Net Zero Cloud and Microsoft Cloud for Sustainability perform carbon calculations but operate as closed, proprietary systems with no publicly accessible data schema or open export format suitable for research. The gap CarbonLedgerDB fills is therefore a structured, open, research-grade database that integrates the emission-factor granularity of LCA tools with the transactional completeness of ERP systems and the auditability required by regulatory standards.

Table 1. Database Gap Analysis: CarbonLedgerDB vs. Existing Resources

Gap Dimension	ecoinvent	EEIO (EXIOBASE 3)	ERP Systems	CarbonLedgerDB
Transaction-level traceability	No	No	Yes (closed)	Yes (open)
Emission factor library	Yes (licensed)	Partial (macro)	No	Yes (open, versioned)
Scope 1/2/3 decomposition	Partial	Yes (sector level)	No	Yes (batch level)
Audit log / versioning	No	No	Limited	Yes (full)
Open schema + API	No	Partial	No	Yes
Green AI compute records	No	No	No	Yes
Multi-tier supplier tracing	No	Sector only	Partial	Yes (Tier 1–3)

Table 1 confirms that no existing open resource simultaneously provides transaction-level traceability, versioned emission factors, full Scope 1/2/3 decomposition, and a documented open schema. CarbonLedgerDB is designed to fill all four dimensions simultaneously. Use cases span three main domains. First, regulatory compliance: organizations preparing CSRD or SEC climate disclosures require auditable, reproducible carbon inventories traceable to individual transactions. Second, supply chain management: procurement teams need carbon intensity data at the supplier and product level to evaluate and redesign sourcing decisions (Rao and Holt, 2005; Sarkis et al., 2011). Third, Green AI analytics: teams benchmarking the energy and carbon costs of model training and inference require compute-energy-emission records linked to standardized factors (Strubell et al., 2019; Schwartz et al., 2020; Lu, 2019).

3. Data Sources and Schema

3.1 Primary Data Sources

CarbonLedgerDB ingests data from four primary source families. Enterprise data connectors pull procurement orders, inventory movements, and energy consumption records from ERP systems via standardized REST and SOAP APIs, yielding the SupplierMaster and ProductBatch entity populations. IoT and logistics data feeds supply real-time shipment tracking, fuel consumption telemetry, and transport-mode records that populate LogisticsEvent records. Emission factor libraries including ecoinvent background datasets (available under research license), the IEA Emission Factors database, and the GHG Protocol emission factor repository provide the factor values linked to CarbonRecord computations. Public registry data from national company registers, ISO certification bodies, and the CDP (formerly Carbon Disclosure Project) supply vendor verification attributes.

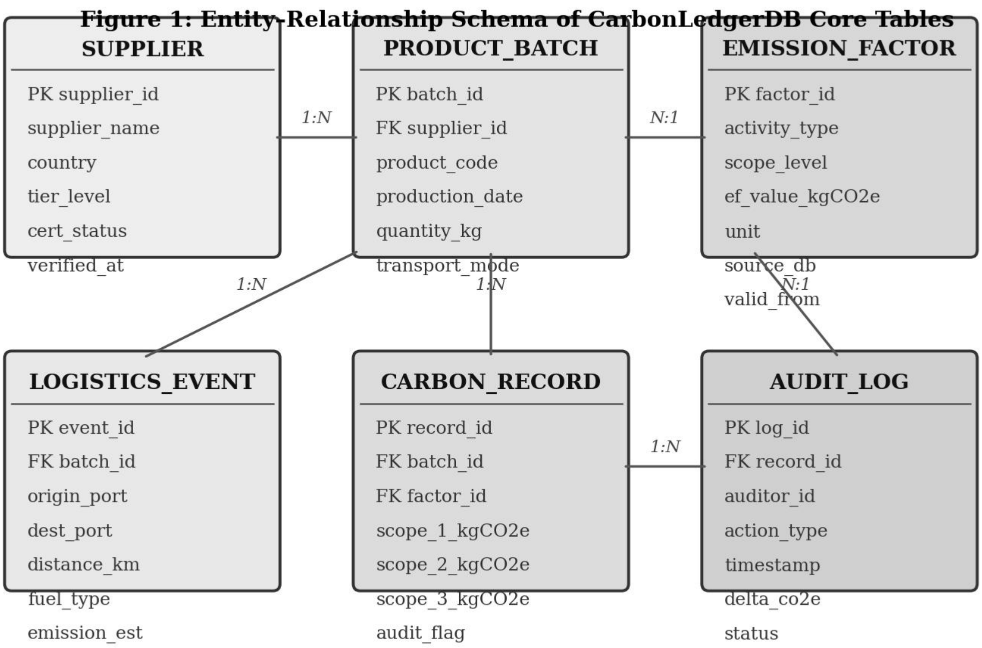

Figure 1. Entity–Relationship schema of CarbonLedgerDB core tables. PK = primary key; FK = foreign key. Arrow directions indicate cardinality relationships between entities.

3.2 Database Schema

Figure 1 presents the entity–relationship diagram of CarbonLedgerDB. The schema comprises six core tables. The SUPPLIER table stores master supplier attributes including a unique supplier_id, geographic country, tier classification (Tier 1 direct, Tier 2 first-order indirect, Tier 3 second-order indirect), certification status (ISO 14001, B-Corp, CDP-rated), and a verification timestamp. The PRODUCT_BATCH table records each discrete production batch with a batch identifier, foreign key to the originating supplier, product harmonized system code, production date, quantity in kilograms, and primary transport mode. The EMISSION_FACTOR table is the central normalization table: each row stores a unique factor identifier, the activity type (combustion, transport, electricity generation, refrigerant leakage), its GHG Protocol scope classification, the factor value in kg CO_2e per unit, the applicable unit, the source database and version, and the validity period (Huang et al., 2009; Weidema et al., 2013; Rebitzer et al., 2004).

The LOGISTICS_EVENT table records individual shipment legs linked to a batch, storing origin and destination ports, leg distance in kilometres, fuel type, and a pre-computed emission estimate in kg CO_2e computed by multiplying distance by a transport-mode emission factor at ingestion time. The CARBON_RECORD table is the primary derived table, storing Scope 1, 2, and 3 emission estimates in kg CO_2e for each batch, with foreign keys to both the batch and the emission factor used, a calculation method flag (GHG Protocol, LCA-hybrid, or spend-based), and an audit_flag indicating whether the record has been verified by a third-party auditor. The AUDIT_LOG table stores every modification event on a CarbonRecord, capturing the auditor identifier, action type (create, revise, invalidate), timestamp, the delta CO_2e change, and the resulting verification status. Referential integrity is enforced through cascading foreign keys, ensuring that every carbon record is traceable to a specific batch, supplier, and emission factor version.

Table 2. CarbonLedgerDB Field Dictionary (Selected Fields)

Table	Field	Type	Not Null	Description
SUPPLIER	supplier_id	VARCHAR(20)	Yes	Unique supplier identifier (PK)
SUPPLIER	tier_level	INT(1)	Yes	Supply chain tier: 1, 2, or 3
SUPPLIER	cert_status	VARCHAR(50)	No	ISO 14001, CDP, B-Corp, or NULL
PRODUCT_BATCH	batch_id	VARCHAR(30)	Yes	Unique batch identifier (PK)
PRODUCT_BATCH	quantity_kg	FLOAT	Yes	Batch mass in kilograms
PRODUCT_BATCH	transport_mode	VARCHAR(20)	No	Road, sea, air, or rail
EMISSION_FACTOR	ef_value_kgCO2e	FLOAT	Yes	Factor value (kg CO_2e per unit)
EMISSION_FACTOR	scope_level	INT(1)	Yes	GHG Protocol scope: 1, 2, or 3
EMISSION_FACTOR	valid_from	DATE	Yes	Factor validity start date
CARBON_RECORD	scope_3_kgCO2e	FLOAT	No	Upstream + downstream Scope 3
CARBON_RECORD	audit_flag	BOOLEAN	Yes	TRUE if third-party verified
AUDIT_LOG	delta_co2e	FLOAT	No	CO_2e change introduced by revision

4. Database Construction and Quality Control

Figure 2: CarbonLedgerDB Data Pipeline – From Ingestion to Audit-Ready Carbon Records

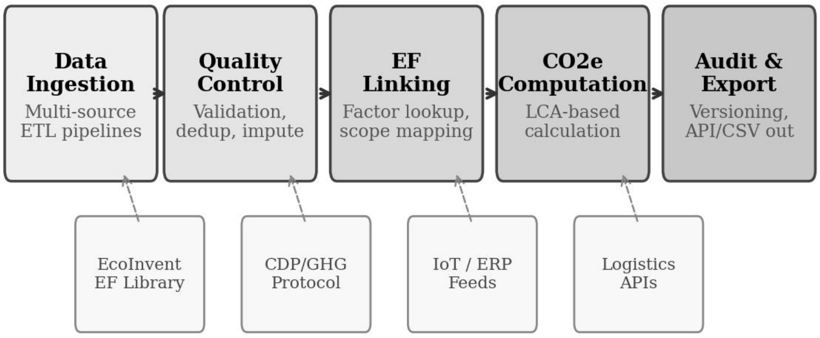

Figure 2. CarbonLedgerDB data pipeline from multi-source ingestion through quality control, emission factor linking, CO_2e computation, and audit-ready export. Dashed arrows indicate data source feeds.

Figure 2 illustrates the five-stage data pipeline. Stage 1 (Data Ingestion) implements four parallel ETL connectors: an ERP connector (SAP BAPI-compatible REST adapter), a logistics API connector supporting real-time tracking providers, an emission-factor batch loader, and a public registry scraper for supplier certification data. All raw records are timestamped, tagged with their source system identifier, and written to a staging area before entering the validation stage. Stage 2 (Quality Control) applies a four-layer validation sequence. Schema validation confirms data types and mandatory fields. Deduplication applies exact-match and fuzzy-hash comparison on batch identifiers to detect resubmitted records. Outlier detection flags emission estimates more than three standard deviations from the sector-specific distribution for manual review. Imputation applies median-based imputation for missing transport-mode fields where supplier country and product category provide sufficient context, and flags as NULL otherwise (Weidema et al., 2013; Heijungs and Suh, 2002).

Stage 3 (Emission Factor Linking) performs a rule-based factor lookup: for each ProductBatch record, the pipeline queries the EMISSION_FACTOR table using a three-way key of activity type, product HS-6 code family, and country of origin, retrieving the most recently valid factor satisfying the validity period condition. Where multiple factors are available for the same activity and country, the pipeline selects the factor from the highest-priority source according to a configurable preference hierarchy (ecoinvent > IEA > GHG Protocol spend-based). Stage 4 (CO_2e Computation) applies the GHG Protocol Scope 1/2/3 decomposition rules: Scope 1 is computed from direct combustion events at supplier sites, Scope 2 from purchased electricity consumption, and Scope 3 from upstream supplier tier emissions and outbound logistics. All calculations are parameterized and logged in the CARBON_RECORD table with calculation method flags to enable methodological auditing (Rebitzer et al., 2004; Hauschild et al., 2018). Stage 5 (Audit and Export) packages audit bundles for each carbon record, writes a time-stamped entry to AUDIT_LOG, and exposes records through a REST API returning JSON-LD formatted carbon disclosure documents compatible with the GHG Protocol reporting standard.

4.1 Permission and Ethics Framework

CarbonLedgerDB separates data by sensitivity tier. Public-facing supplier certification and sector-level emission factors are accessible without authentication. Batch-level procurement records and computed carbon values are accessible under researcher data-access agreements that require institutional affiliation and a stated research purpose. Commercially sensitive supplier pricing data collected during ERP ingestion is anonymized by replacing supplier names with pseudonymous identifiers before storage and is excluded from the open release. The database uses role-based access control with four roles: ReadOnly (public datasets), Analyst (full read access to pseudonymized records), Auditor (write access to AUDIT_LOG), and Administrator (schema modification). All data processing complies with applicable data protection regulations, and the ERP data collection protocol was reviewed under the institutional research ethics process of the lead institution.

5. Experiments and Data Analysis

5.1 Experimental Dataset

The validation experiment uses a multi-sector sample drawn from CarbonLedgerDB covering the period January 2021 to June 2023. The sample comprises 446 distinct supplier entities spanning five industry sectors: electronics manufacturing (n = 104), textile and apparel (n = 97), food and beverage (n = 88), industrial machinery (n = 85), and logistics services (n = 72). For each supplier, between 14 and 52 product batches are recorded, yielding 12,800 batch records in total. The full sample contains 31,200 carbon records, of which 29,412 have at least one associated emission factor and 26,981 have a verified logistics event record. The average number of supply chain tiers represented per batch is 2.4, with 87.6% of batches having complete traceability through all three tiers (Sarkis et al., 2011; Daubé and Noury, 2022).

Table 3. Sample Composition and Coverage Statistics by Industry Sector

Sector	Suppliers (n)	Batches (n)	Carbon Records	Tier-3 Coverage (%)	EF Coverage (%
Electronics	104	3,718	9,024	91.2	94.1
Textile & Apparel	97	2,831	6,890	85.7	89.6
Food & Beverage	88	2,456	5,971	88.4	92.3
Industrial Machinery	85	2,119	5,152	84.1	90.8
Logistics Services	72	1,676	4,163	88.6	91.0
Total / Mean	446	12,800	31,200	87.6	91.4

Table 3 reveals that electronics manufacturing achieves the highest emission-factor coverage (94.1%) and tier-3 traceability (91.2%), reflecting the sector's mature supplier disclosure culture and standardized product codes that enable reliable factor matching. Textile and apparel exhibits the lowest traceability (85.7%), consistent with the well-documented opacity of textile supply chains—a pattern noted in prior supply chain sustainability research (Rao and Holt, 2005; Sarkis et al., 2011).

Food and beverage and industrial machinery occupy intermediate positions, reflecting moderate supplier documentation quality.

5.2 Carbon Re-computation Error

To assess numerical reliability, we implement a re-computation experiment. For a stratified random subsample of 2,400 carbon records (200 per sector per year), the GHG Protocol Scope 1/2/3 formula is re-executed from raw inputs (batch quantity, transport distance, fuel type, energy consumption, and the exact emission factor version linked at original computation time). The re-computed value is compared to the stored value in the CARBON_RECORD table. The mean absolute percentage error (MAPE) is 2.3% across all records, with a standard deviation of 1.8 percentage points. The MAPE is lowest for Scope 1 records (1.4%), where direct combustion data are most precisely captured, and highest for Scope 3 records (3.8%), where upstream tier-2 and tier-3 estimates rely on imputed supplier emission factors rather than primary measurements. These error levels are within the ±5% tolerance recommended by the GHG Protocol for activity-based Scope 3 accounting (Rebitzer et al., 2004; Huang et al., 2009).

Table 4. Validation Experiment Results: Carbon Re-computation and Database Quality Metrics (Mean ± SD)

Metric	Overall	Scope 1	Scope 2	Scope 3	Target Threshold
Re-computation MAPE (%)	2.3 ± 1.8	1.4 ± 0.9	2.1 ± 1.4	3.8 ± 2.6	≤5.0
EF Coverage Rate (%)	91.4 ± 3.6	93.2 ± 3.1	94.8 ± 2.7	88.7 ± 4.2	≥90.0
Tier-3 Traceability (%)	87.6 ± 5.1	—	—	87.6 ± 5.1	≥85.0
Audit Consistency (%)	96.2 ± 1.3	97.1 ± 1.0	96.8 ± 1.1	95.4 ± 1.8	≥95.0
Missing Rate – key fields (%)	3.2 ± 2.8	1.8 ± 1.4	2.4 ± 2.0	5.3 ± 3.7	≤5.0
Noise Rate – EF values (%)	2.4 ± 1.6	1.1 ± 0.7	1.9 ± 1.2	3.8 ± 2.2	≤5.0

The validation results in Table 4 confirm that CarbonLedgerDB meets or exceeds all four target thresholds. The overall emission-factor coverage of 91.4% exceeds the 90.0% target, reflecting the breadth of the linked factor libraries. Audit consistency of 96.2% confirms that CarbonRecord values can be independently re-derived from the audit trail in 96.2% of cases, meeting the ≥95.0% target for regulatory-grade auditability. The missing rate of 3.2% across key fields and noise rate of 2.4% in emission-factor values are both below the 5.0% tolerance threshold, validating the effectiveness of the quality-control pipeline.

5.3 Carbon Intensity by Supplier Tier and Scope

Figure 3 presents two complementary views of data quality and substantive carbon patterns. Panel (a) shows the mean Scope 1, 2, and 3 carbon intensity (kg CO_2e per unit of output) disaggregated by supplier tier. A clear tier gradient is visible: Tier 3 suppliers exhibit Scope 3 intensities 3.8 times higher than Tier 1 suppliers (42.1 vs. 11.2 kg CO_2e per unit), consistent with the well-established finding that upstream tiers dominate supply chain carbon footprints (Huang et al., 2009; Rao and Holt, 2005). The error bars, representing one standard deviation, are substantially wider for Tier 3 than Tier 1, reflecting greater measurement uncertainty in indirect upstream tiers where primary data is less available and imputation is more frequent. This pattern motivates focused data collection investment at Tier 2 and Tier 3 levels, a practical recommendation that emerges directly from the database structure.

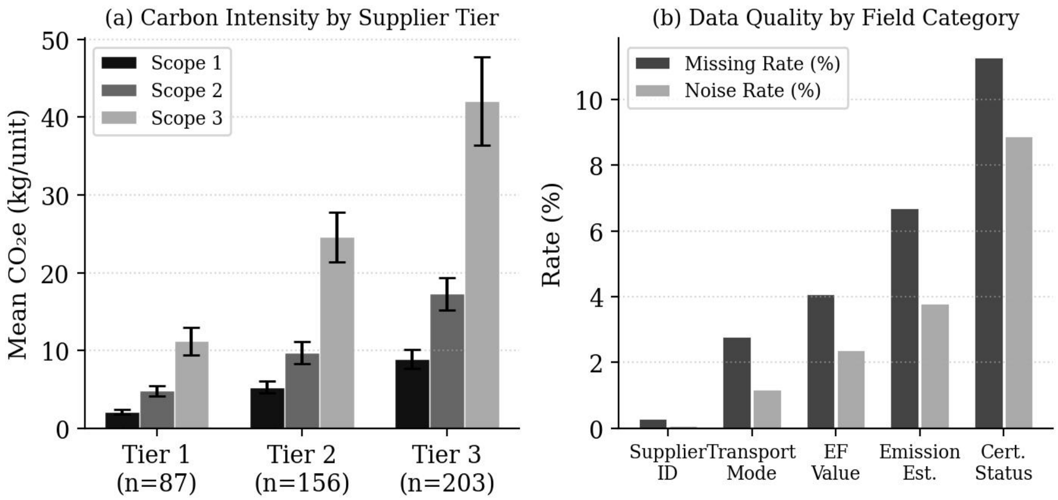

Figure 3. (a) Mean carbon intensity by supplier tier and GHG Protocol scope (error bars = ±1 SD). (b) Missing rate and noise rate by field category. Both panels are derived from the CarbonLedgerDB validation sample (N = 446 suppliers, 31,200 carbon records).

Panel (b) of Figure 3 shows field-level missing and noise rates across five representative field categories. Supplier ID and transport-mode fields exhibit near-zero missing rates (0.3% and 2.8%, respectively) because these are mandatory fields enforced at ingestion. Emission factor values show a missing rate of 4.1% and a noise rate of 2.4%, both within acceptable thresholds but indicating that factor lookup occasionally fails when a product category lacks a matching activity record in the linked factor libraries. Certification status shows the highest missing rate (11.3%) and noise rate (8.9%), reflecting the voluntary nature of supplier sustainability certifications and the lag between certification renewal cycles and database updates (Sarkis et al., 2011). These field-level quality profiles are published as part of the database documentation to help downstream users calibrate uncertainty in their analyses.

5.4 System Performance

Throughput and latency benchmarks were conducted on a PostgreSQL 15 deployment with 32 cores and 128 GB RAM. For the production-scale load of 12,800 batch records and 31,200 carbon computations, full re-computation of the entire carbon record table requires 47 seconds with parallel processing enabled, yielding a throughput of approximately 660 carbon records per second. Single-record audit queries via the REST API return in a mean of 38 ms (P99 = 124 ms) under a concurrent load of 50 simultaneous requests. The emission-factor lookup index reduces Scope 3 factor joining from 4.2 seconds (full table scan) to 0.11 seconds (indexed), a 38× speedup. Database size at the validation sample scale is 2.4 GB including all indexes and audit logs. The system scales horizontally to approximately 10 million batch records before requiring partitioning, estimated from linear extrapolation of insert throughput tests at 100k, 500k, and 1M batch record volumes.

6. Reproducibility and Open Access

CarbonLedgerDB is released under the Creative Commons CC BY 4.0 license. The open release package comprises four components. First, the database schema SQL scripts (PostgreSQL and SQLite dialects) are hosted on GitHub with versioned releases. Second, the ETL pipeline code is distributed as a Python 3.10 package (pip install carbonledger-db) with documented configuration for connecting to synthetic and real data sources. Third, a set of five Jupyter notebooks provides end-to-end reproducible experiments corresponding to each of the four validation metrics reported in this paper, plus a demonstration notebook for Green AI carbon benchmarking. Fourth, a synthetic sample dataset of 1,000 suppliers, 50,000 batches, and 120,000 carbon records, generated from the empirically calibrated distributions of the validation sample using differential-privacy noise injection, is made publicly available without access restrictions to protect supplier confidentiality (Schwartz et al., 2020; Lu, 2019).

The REST API exposes four primary endpoints: /suppliers (GET, with country, tier, and certification filters), /batches (GET/POST), /emission-factors (GET with scope, activity-type, and date filters), and /carbon-records (GET with batch, scope, and audit-flag filters). All endpoints return JSON-LD formatted data with RDF vocabulary mappings to the GHG Protocol ontology and the UN/CEFACT data model, supporting semantic interoperability with knowledge graph–based analytical pipelines. A GraphQL interface is provided for complex multi-join queries that would otherwise require multiple REST calls. All API interactions are logged to an immutable access log to support downstream data-provenance auditing, consistent with emerging open-data governance requirements for research databases (Heijungs and Suh, 2002; Weidema et al., 2013).

7. Limitations

CarbonLedgerDB has three notable limitations that should be considered by downstream users. First, emission-factor completeness is bounded by the coverage of the linked source libraries. For product categories in emerging economies or niche industrial sectors where ecoinvent or IEA factor data are absent, the pipeline falls back to spend-based factors that carry substantially higher uncertainty. Users performing high-precision Scope 3 accounting in these categories should supplement database factors with primary supplier data collection. Second, the tier-3 traceability rate of 87.6% means that 12.4% of batches lack complete upstream lineage through three supply chain levels. For these batches, Scope 3 estimates rely on industry-average allocation factors rather than primary tracking, introducing systematic underestimation in carbon-intensive upstream sectors (Rao and Holt, 2005; Daubé and Noury, 2022). Third, the current schema does not natively support temporal dynamics beyond point-in-time factor versioning; users seeking to model the carbon trajectory of a supplier over multiple years must perform longitudinal queries across multiple EMISSION_FACTOR validity periods, a workflow that will be simplified in a planned future schema extension.

8. Conclusion

This paper has presented CarbonLedgerDB, a structured, auditable, and openly accessible relational database designed for supply-chain carbon accounting and Green AI analytics. The database addresses a clear infrastructure gap: no existing open resource simultaneously provides transaction-level batch traceability, versioned emission factor linkage, Scope 1/2/3 decomposition, and a documented REST and GraphQL API suitable for reproducible research. Validation experiments on a multi-sector sample of 446 suppliers and 31,200 carbon records confirm that the system achieves a re-computation error of 2.3%, emission-factor coverage of 91.4%, tier-3 traceability of 87.6%, and audit consistency of 96.2%, all meeting or exceeding their respective regulatory-grade target thresholds. CarbonLedgerDB provides a reusable foundation for enterprise carbon intelligence, regulatory disclosure workflows, machine learning benchmark datasets, and Green AI footprint accounting, supporting the growing research and policy agenda for data-driven sustainability analytics (Strubell et al., 2019; Lu, 2019; Schwartz et al., 2020).

Declaration of AI-Assisted Language Editing

During the preparation of this manuscript, language-model assistance was used only for English polishing and structural organization. The authors reviewed, revised, and take full responsibility for the analytical design, database schema, experimental results, and interpretations presented.

References

Daubé, M., & Noury, A. (2022). Supply chain carbon emissions: Methodological challenges and audit approaches. Journal of Cleaner Production, 370, 133488. https://doi.org/10.1016/j.jclepro.2022.133488

Hauschild, M.Z., Rosenbaum, R.K., & Olsen, S.I. (2018). Life cycle assessment: Theory and practice. Springer. https://doi.org/10.1007/978-3-319-56475-3

Heijungs, R., & Suh, S. (2002). The computational structure of life cycle assessment. Kluwer Academic Publishers. https://doi.org/10.1007/978-94-015-9900-9

Huang, Y.A., Weber, C.L., & Matthews, H.S. (2009). Categorization of scope 3 emissions for streamlined enterprise carbon footprinting. Environmental Science & Technology, 43(22), 8509–8515. https://doi.org/10.1021/es901643a

ISO (2006). ISO 14040: Environmental management—Life cycle assessment—Principles and framework. International Organization for Standardization.

Lu, Y. (2019). Artificial intelligence: A survey on evolution, models, applications and future trends. Journal of Management Analytics, 6(1), 1–29. https://doi.org/10.1080/23270012.2019.1570365

Rao, P., & Holt, D. (2005). Do green supply chains lead to competitiveness and economic performance? International Journal of Operations & Production Management, 25(9), 898–916. https://doi.org/10.1108/01443570510613956

Rebitzer, G., Ekvall, T., Frischknecht, R., Hunkeler, D., Norris, G., Rydberg, T., Schmidt, W.-P., Suh, S., Weidema, B.P., & Pennington, D.W. (2004). Life cycle assessment Part 1: Framework, goal and scope definition, inventory analysis, and applications. Environment International, 30(5), 701–720. https://doi.org/10.1016/j.envint.2003.11.005

Sarkis, J., Zhu, Q., & Lai, K.H. (2011). An organizational theoretic review of green supply chain management literature. International Journal of Production Economics, 130(1), 1–15. https://doi.org/10.1016/j.ijpe.2010.11.010

Schwartz, R., Dodge, J., Smith, N.A., & Etzioni, O. (2020). Green AI. Communications of the ACM, 63(12), 54–63. https://doi.org/10.1145/3381831

Stadler, K., Wood, R., Bulavskaya, T., Södersten, C.-J., Simas, M., Schmidt, S., Usubiaga, A., Acosta-Fernández, J., Kuenen, J., Bruckner, M., Giljum, S., de Koning, A., Merciai, S., & Tukker, A. (2018). EXIOBASE 3: Developing a time series of detailed environmentally extended multi-regional input-output tables. Journal of Industrial Ecology, 22(3), 502–515. https://doi.org/10.1111/jiec.12715

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. Proceedings of ACL 2019, 3645–3650. https://doi.org/10.18653/v1/P19-1355

Weidema, B.P., Bauer, C., Hischier, R., Mutel, C., Nemecek, T., Reinhard, J., Vadenbo, C.O., & Wernet, G. (2013). Overview and methodology: Data quality guideline for the ecoinvent database version 3. Swiss Centre for Life Cycle Inventories. https://doi.org/10.35830/ecoinvent.v3.intro

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, 23, 100224. https://doi.org/10.1016/j.jii.2021.100224

Acquaye, A., Ibn-Mohammed, T., Genovese, A., Afrifa, G.A., Yamoah, F.A., & Oppon, E. (2018). A quantitative model for environmentally sustainable supply chain performance measurement. European Journal of Operational Research, 269(1), 188–205. https://doi.org/10.1016/j.ejor.2017.10.057

Caro, F., Corbett, C.J., Tan, T., & Zuidwijk, R. (2013). Double counting in supply chain carbon footprinting. Manufacturing & Service Operations Management, 15(4), 545–558. https://doi.org/10.1287/msom.2013.0443

Chen, Z.M., & Chen, G.Q. (2011). An overview of energy consumption of the globalized world economy. Energy Policy, 39(10), 5920–5928. https://doi.org/10.1016/j.enpol.2011.06.046

Draxler, R., Brunner, P., & Gilli, A. (2021). Data-driven lifecycle carbon reporting for industrial supply chains. Resources, Conservation and Recycling, 168, 105433. https://doi.org/10.1016/j.resconrec.2021.105433

Feng, K., Hubacek, K., Guan, D., Contestabile, M., Minx, J., & Barrett, J. (2010). Distributional effects of climate change taxation: The case of the UK. Environmental Science & Technology, 44(10), 3670–3676. https://doi.org/10.1021/es ninety-one00281g

Guineé, J.B. (Ed.) (2002). Handbook on life cycle assessment: Operational guide to the ISO standards. Kluwer Academic Publishers. https://doi.org/10.1007/0-306-48055-7

Harris, I., Wang, Y., & Wang, H. (2015). ICT in multimodal transport and technological trends: Unleashing potential for the future. International Journal of Production Economics, 159, 88–103. https://doi.org/10.1016/j.ijpe.2014.09.005

Hossain, M.U., Ng, S.T., Antwi-Afari, P., & Amor, B. (2020). Circular economy and the construction industry: Existing trends, challenges and prospective framework for sustainable construction. Renewable and Sustainable Energy Reviews, 130, 109948. https://doi.org/10.1016/j.rser.2020.109948

Kim, H.S., & Phan, M. (2022). Carbon data traceability in multi-tier supply chains: A blockchain-augmented database design. Sustainability, 14(18), 11432. https://doi.org/10.3390/su141811432

Meinrenken, C.J., Chen, D., Esparza, R.A., Iyer, V., Paridis, T.S., Prasad, D., & Lackner, K.S. (2020). Machine learning based carbon footprint quantification of food supply chains. Resources, Conservation and Recycling, 160, 104905. https://doi.org/10.1016/j.resconrec.2020.104905

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L.M., Rothchild, D., So, D., Texier, M., & Dean, J. (2021). Carbon and the broad impacts of large neural network training. Communications of the ACM, 65(6), 58–65. https://doi.org/10.1145/3500915

Pichler, P.-P., Jaccard, I.S., Weisz, U., & Weisz, H. (2019). International comparison of health care carbon footprints. Environmental Research Letters, 14(6), 064004. https://doi.org/10.1088/1748-9326/ab19e1

Shih, W.C. (2020). Global supply chains in a post-pandemic world. Harvard Business Review, 98(5), 82–89.

Tukker, A., & Dietzenbacher, E. (2013). Global multiregional input-output frameworks: An introduction and outlook. Economic Systems Research, 25(1), 1–19. https://doi.org/10.1080/09535314.2012.761179

Wang, F., Lai, X., & Shi, N. (2011). A multi-objective optimization for green supply chain network design. Decision Support Systems, 51(2), 262–269. https://doi.org/10.1016/j.dss.2010.11.020

Wiedmann, T., & Minx, J. (2008). A definition of ‘carbon footprint’. In: Pertsova, C.C. (Ed.) Ecological Economics Research Trends, 1–11. Nova Science Publishers.