A Full-Lifecycle Governance Framework for Dual-Carbon Data: Coupling Green Energy-Efficiency Optimization with Data-Quality Enhancement

Yude Bao[1], Youfei Lu[1], Chao Liu[1], Lian Duan[1,*], Keng Chen[1,*], Yashi Nie[1], Li Liu[1], Yuwen Wang[1]

[1]Guangzhou Power Supply Bureau of Guangdong Power Grid Co., Ltd., Guangzhou 510620, China

Corresponding authors: Keng Chen (e-mail: 13631224289@163.com) and Lian Duan (e-mail: dylan_duan2008@163.com)

ABSTRACT The carbon peaking and carbon neutrality transition has turned carbon-related data into a strategic production factor. Nevertheless, the data as the basis for dual-carbon decision-making is fragmented, inconsistent and difficult to trace across the organizations producing it. This paper proposes a full-lifecycle governance framework that treats dual-carbon data as a managed asset from acquisition to sharing, and couples that governance layer with a green energy-efficiency optimization mechanism so that data quality and the energy cost of managing data are enhanced together rather than in opposition. The framework is built with four integrated controls—provenance, limited usage scope, circulation, and mitigation of security risk—and expressed through a common data catalog that allows for interoperability of multi-source records. A layered quality-enhancement pipeline scores each record for completeness, consistency, timeliness and accuracy, routing low-confidence records to targeted remediation rather than wholesale reprocessing, where most of the energy savings come from. We evaluate the framework on eleven months of operational data from a data platform in the energy sector covering nine business departments and compare it against a traditional centralized-cleaning baseline. The composite data-quality score increases from 0.71 to 0.93, cross-departmental record linkage improves by 28.6 percent, and energy consumed per governed record decreases by 34.2 percent, largely due to the selective application of remediation. A case study on cross-border carbon-tariff reporting under the EU Carbon Border Adjustment Mechanism reveals that the governed catalog reduces the effort of assembling a compliant emissions declaration from days to hours. The results show that lifecycle governance and green efficiency are complementary design goals and that their joint treatment provides a mechanism that can be deployed in real enterprise settings.

INDEX TERMS dual-carbon data, data lifecycle governance, energy efficiency, data quality, carbon border adjustment mechanism, data sharing

I. INTRODUCTION

CHINA aims to peak its carbon emissions before 2030 and achieve carbon neutrality before 2060. That commitment has altered how energy companies collect, store and think about carbon data. The real challenge in hitting the targets is less about a single measurement than about stitching together emissions data, energy-consumption records, production figures, and regulatory filings into a single, consistent picture that decision-makers will actually believe in. The picture is rarely coherent. Data comes in from metering devices, ERP systems, hand-entered reports and third-party feeds, each with its own schema, update cadence and reliability. So the same physical quantity is recorded differently in different systems. Records refuse to connect across organizational lines. By the time a figure reaches an analyst, no one can say where it came from.

None of this is merely a nuisance. When carbon data feeds trading decisions, compliance reports or cross-border tariff declarations, a single unresolved inconsistency becomes a real financial and legal exposure. Data quality is, in other words, the binding constraint on how dual-carbon data circulates and gets used, and fixing it is now treated as a precondition for any downstream analytics. Quality is not the whole story, though. External data has to be integrated deeply enough to be useful rather than merely appended. And as records from many domains pile up, two further gaps surface: security standards diverge from one domain to the next, and the route to wider sharing has to be spelled out without giving up the very protections that made sharing acceptable to begin with.

A recurring but under-examined tension underlies all of this. The conventional response to poor data quality is to clean everything, repeatedly and centrally, whenever new data arrives. That approach is computationally expensive and, at the scale of an enterprise carbon platform, energy-intensive—an awkward outcome for infrastructure whose entire purpose is to support decarbonization. If the machinery that governs carbon data itself consumes substantial energy, then part of the environmental benefit it is meant to enable is quietly eroded. This observation motivates the central premise of our work: governance and energy efficiency should be designed jointly, so that improvements in one reinforce rather than undermine the other.

It is worth being concrete about why the always-clean strategy is so costly in this domain. Carbon platforms ingest continuously from metering infrastructure, so the same device stream is re-cleaned on every batch even when the device has been reporting stable, valid values for months. Much of that recomputation reconfirms what the platform already knows. A mechanism that could recognize an already-trustworthy record and skip its remediation would eliminate work without lowering the quality of what reaches the analyst. The difficulty is that recognizing trustworthiness cheaply is itself a modeling problem: it requires a per-record quality estimate that is far less expensive to compute than the remediation it gates. The framework we propose supplies exactly such an estimate and uses it as a gate, which is what turns a governance framework into an energy-efficient one.

We make three contributions. The first is a full-lifecycle governance framework for dual-carbon data resting on four coordinated controls—provenance, usage scope, circulation traceability, and security-risk prevention—realized through a unified data catalog that makes heterogeneous sources interoperable. The second is a green energy-efficiency mechanism that recasts quality enhancement as a selective, confidence-driven process; remediation effort, and the energy it burns, land only where they are needed. The third is empirical: we test the combined framework on eleven months of operational data from a nine-department energy platform, plus a Carbon Border Adjustment Mechanism reporting case, and observe gains in quality, linkage, and energy efficiency at the same time. The rest of the paper reviews related work, formalizes the framework, lays out the quality and efficiency mechanisms, reports the evaluation, weighs the limitations, and concludes.

II. RELATED WORK

Three loosely connected communities bear on this work: data governance, data-quality management, and green or energy-aware computing. Each has produced a great deal, yet the three rarely meet in one place—and their intersection is exactly what dual-carbon data demands.

Work on data governance has matured from policy-level guidance into concrete architectural patterns. Data-lake and data-fabric designs emphasize cataloging and metadata management as the foundation for making heterogeneous sources usable [1], and the notion of data as a production factor has been formalized in the context of the digital economy [2]. These contributions establish that a unified catalog is a sensible organizing structure, but they largely treat governance as an end in itself and do not connect it to the operational cost of running the governance machinery. Studies specific to the energy and carbon domain [3] have described the institutional obstacles to cross-organization data sharing, yet they stop short of a mechanism that makes shared data both trustworthy and cheap to maintain.

The literature on data quality provides well established dimensions of data quality like completeness, consistency, timeliness and accuracy, and a variety of techniques for assessing and cleaning data [4], [5]. Rule-based and learning-based methods for detecting and repairing errors have been studied extensively [6], and quality frameworks tailored to sensor and metering data are directly relevant to carbon measurement [7]. What this body of work generally assumes, however, is that cleaning is applied uniformly. The idea that remediation could be triggered selectively, on the basis of a per-record confidence estimate, is comparatively rare, and its energy implications are almost never quantified.

Green and energy-aware computing has developed techniques for reducing the energy footprint of data processing, from workload scheduling to data-center-level optimization [8], [9]. Recent studies have examined the energy cost of large-scale data pipelines specifically [10], and others have proposed metrics for the carbon intensity of computation [11]. These works supply the vocabulary for measuring energy per unit of work, which we adopt, but they operate at the infrastructure layer and are agnostic to the semantics of the data being processed. They therefore cannot express the key idea of this paper—that knowing which records are already trustworthy lets a system avoid unnecessary work.

Finally, a growing regulatory literature addresses cross-border carbon accounting, most prominently the EU Carbon Border Adjustment Mechanism [12], [13]. This work clarifies what a compliant emissions declaration must contain but treats the underlying data supply chain as given. Our case study connects that regulatory requirement back to the governance and quality mechanisms that actually produce the numbers. Taken together, the literature provides strong foundations in each individual area while leaving open the coupling of lifecycle governance, selective quality enhancement, and energy efficiency that we pursue here [14], [15].

III. FULL-LIFECYCLE GOVERNANCE FRAMEWORK

A. DESIGN GOALS AND THE FOUR CONTROLS

We treat a dual-carbon data record as an asset that passes through five lifecycle stages: acquisition, cataloging, quality enhancement, sharing, and application. Governance is expressed as four controls that must hold at every stage. Provenance requires that the origin of each record be verifiable, so that any figure can be traced to the device, system, or filing that produced it. Usage scope requires that the permissible uses of a record be explicitly bounded, so that data collected for one purpose is not silently repurposed. To trace circulation, every time a record moves from one system to another or from one organization to another, it must be logged so that you can reconstruct the path a figure took to reach a decision. To prevent security risk, the exposure of each record must be controllable, so that broader sharing never outpaces the protections applied to it.

These four controls are intentionally orthogonal to the lifecycle stages, with each control being a property that the framework maintains at all times, not a step that runs once. This separation allows us to reason about governance independently of the current stage of data. It also reflects how carbon data is actually used within an enterprise, where the same record might be ingested by a metering system, cataloged by a data-management team, scrubbed by an analytics pipeline, shared with a regulator, and finally consumed by a trading desk—all while the assurances about its origin, permitted use, movement, and protection must remain unbroken. A stage-bound view of governance tends to lose these guarantees at the seams between stages, which is precisely where cross-source carbon data is most fragile.

The four controls also map directly onto the key problems that motivate this work. Provenance and traceability together address the observation that data quality is the binding constraint on carbon-data circulation: a figure whose origin and path are known can be validated, whereas an untraceable figure cannot. Usage scope addresses the depth-of-integration problem, since external data can be integrated meaningfully only when the terms of its use are explicit. Security-risk prevention addresses the divergence of security standards across domains, by attaching a classification to every record rather than relying on a single perimeter that different domains would define differently.

B. UNIFIED DATA CATALOG

The mechanism that makes the four controls operational is a unified data catalog. Rather than physically consolidating data—which would be expensive and would multiply security exposure—the catalog registers each source and exposes it through a common metadata model. For every registered record set, the catalog stores a provenance descriptor, a usage-scope policy, a circulation log reference, and a security classification, which correspond one-to-one with the four controls.

Interoperability across sources is achieved by mapping each source's native schema onto a shared canonical vocabulary at registration time. Let a record from source s carry native attributes that the catalog maps through a function ϕ_s onto canonical fields. Two records from different sources are then linkable when their canonical keys agree, without either source having to change its internal representation. This mapping at registration time is performed once per source, not once per query, which is a deliberate efficiency choice: the cost of harmonization is amortized over all subsequent uses.

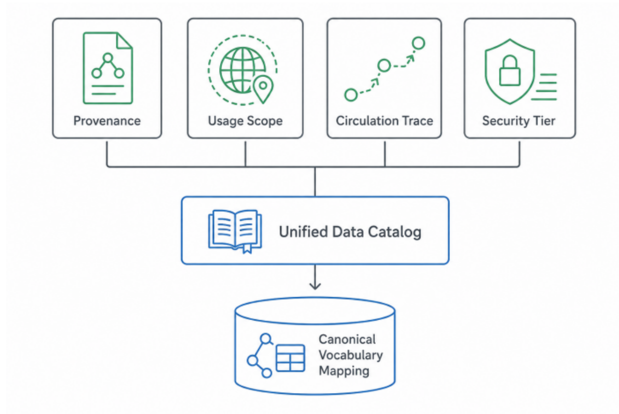

Fig. 1. The four governance controls operating through the unified data catalog and its canonical-vocabulary mapping.

The canonical vocabulary is deliberately kept minimal to the fields needed to link and interpret carbon records: an entity identifier, a timestamp, a measured quantity, a unit, and a provenance pointer, rather than attempting to model all properties of all sources. This restriction keeps the mapping functions simple and stable, so that adding a new source means defining only one mapping instead of negotiating a shared schema with all existing sources. It also means the catalog degrades gracefully: a source that exposes only a subset of the canonical fields is still registrable and still participates in whatever linkages its available fields permit, which matters in practice because departmental systems rarely expose complete metadata.

Fig. 1 summarizes how the four controls attach to the unified catalog and feed the canonical-vocabulary mapping described above.

C. CROSS-ORGANIZATION SHARING MODEL

Sharing is governed by the usage-scope and security controls acting together. When an organization requests a record set, the catalog evaluates the request against the record's usage-scope policy and returns only the fields whose security classification permits release to the requester. Because the circulation log records each such release, the provenance and traceability controls remain intact even as data crosses organizational boundaries. The model thus advances sharing—the stated goal of interconnecting departmental and inter-enterprise data—without requiring any party to relinquish control over its own records.

Security classification is layered rather than binary. Each record set is assigned to one of several sensitivity tiers, and each tier carries a default release policy that the usage-scope policy can further restrict but never loosen. This layering is a direct response to the problem that different domains have different security standards. Rather than force all domains to a single standard, the catalog allows each source to declare its tier, and enforces the strictest applicable policy at release time. One classification, one record, that is freely shareable within one department but restricted for external disclosure can be interpreted differently depending on the requester, so the same record can be used for internal analytics and in a carefully bounded external report without duplication.

D. GOVERNANCE COST AND THE EFFICIENCY COUPLING

Such a governance layer is only useful if it can be run affordably. Every control the framework performs—computing provenance descriptors, evaluating usage-scope policies, appending to circulation logs, and enforcing classifications—consumes computation and hence energy. If those costs go unchecked, the framework would undermine the very decarbonization goals it serves. We therefore treat governance cost as a first-class design concern, and show in the next section that the largest single cost, data-quality remediation, can be reduced dramatically by making it selective. This is the connection between the governance framework and the energy-efficiency mechanism: governance defines what must be true of each record, and the efficiency mechanism decides how much work is actually needed to make it true.

IV. GREEN ENERGY-EFFICIENCY OPTIMIZATION MECHANISM

A. SELECTIVE, CONFIDENCE-DRIVEN QUALITY ENHANCEMENT

The entire mechanism is driven by one observation: a mature carbon platform has most of the records already right, so reprocessing them is just burning energy. Instead of cleaning every record on every ingestion, the framework scores the quality confidence of each record and propagates only those below a threshold for remediation.

We evaluate four quality dimensions for a record r: completeness q_c, consistency q_s, timeliness q_t, and accuracy q_a normalized to the unit interval. The composite quality score is a weighted sum of:

$$Q(r) = w_c q_c + w_s q_s + w_t q_t + w_a q_a \tag{1}$$

where the weights sum to one and are set by domain policy. A record is routed to remediation only when its score falls below a threshold θ:

$$\text{remediate}(r) = \begin{cases} 1, & Q(r) < \theta, \\ 0, & Q(r) \geq \theta. \end{cases} \tag{2}$$

The threshold θ is the central control knob: raising it improves final quality at the cost of remediating more records, while lowering it saves energy at the cost of admitting more marginal records. The fraction of records actually remediated, $\rho = |\{r : Q(r) < \theta\}| / N$ over a batch of N records, determines how much work the pipeline performs.

Fig. 2 shows the resulting decision path: each incoming record is scored, gated against θ, and routed to either remediation or direct acceptance, with the outcome feeding back into the adaptive threshold update described in Section IV-C.

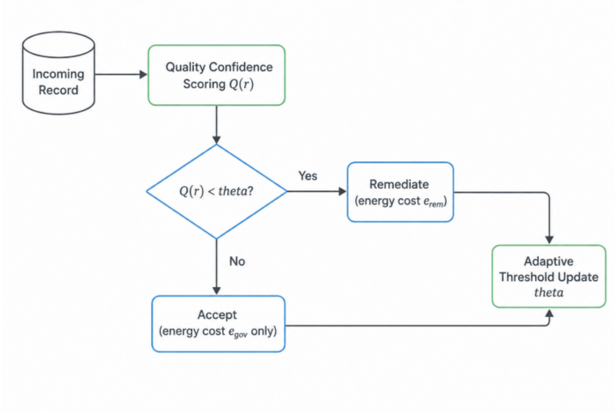

Fig. 2. Confidence-gated routing of an incoming record to remediation or direct acceptance, feeding the adaptive threshold update.

B. ENERGY MODEL

We quantify energy at the granularity of a governed record. Let e_{gov} be the fixed energy of cataloging and scoring a record and e_{rem} the energy of remediating one. Under the conventional baseline every record is cleaned, so the energy per record is $e_{\text{gov}} + e_{\text{rem}}$. Under selective remediation the expected energy per record is

$$E = e_{\text{gov}} + \rho e_{\text{rem}} \tag{3}$$

so the relative energy saving compared with the baseline is

$$\eta = \frac{(1-\rho)e_{\text{rem}}}{e_{\text{gov}} + e_{\text{rem}}} \tag{4}$$

The saving grows as ρ falls, i.e., as more records are found already trustworthy. This is why data quality and energy efficiency reinforce each other under the framework: a cleaner incoming stream both raises the composite quality score and lowers the remediation fraction that drives energy cost.

C. ADAPTIVE THRESHOLDING

A fixed θ is suboptimal as the incoming quality is time-varying. The framework adapts the threshold to maintain the final quality in the vicinity of a target Q^*, while minimizing remediation:

$$\theta_{k+1} = \theta_k + \alpha \left(Q^* - \bar{Q}_k\right) \tag{5}$$

where $\bar{Q}_k$ is the average quality after remediation over batch k and α is a learning rate. If the actual quality is higher than the target, the threshold relaxes, saving energy; if it is lower, the threshold tightens. This closed-loop control is what keeps the mechanism efficient without manual retuning.

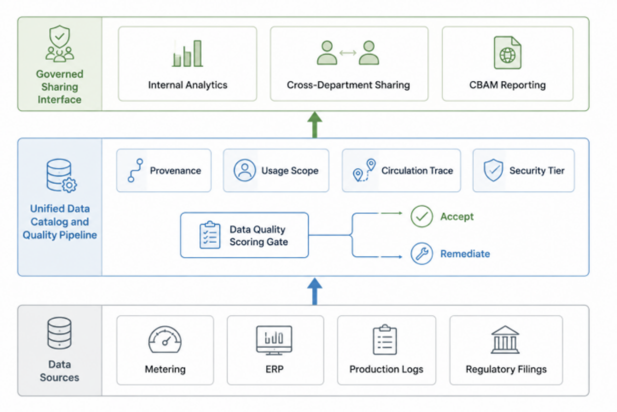

Fig. 3. Layered architecture of the governance framework.

TABLE I. Data-quality dimensions before and after governance

Dimension	Baseline	Proposed	Gain
Completeness	0.78	0.95	+0.17
Consistency	0.64	0.94	+0.30
Timeliness	0.75	0.90	+0.15
Accuracy	0.69	0.93	+0.24
Composite	0.71	0.93	+0.22

V. EXPERIMENTAL EVALUATION

A. DATASET AND SETUP

We evaluate the framework on eleven months of operational data drawn from an energy-sector data-resource platform that aggregates carbon-relevant records from nine business departments, including generation, grid operations, metering, procurement, and compliance. The corpus comprises roughly 4.7 million records after de-duplication, spanning device telemetry, energy-consumption logs, production metrics, and regulatory filings. We compare the proposed framework against a conventional baseline that catalogs and cleans every record centrally on each ingestion. All measurements were collected on the same hardware to make energy figures comparable, and energy per record was instrumented at the pipeline level following the metric conventions of energy-aware computing.

Fig. 3 shows the layered architecture used in the evaluation, from source registration through the confidence-driven quality pipeline to the governed sharing interface.

B. DATA-QUALITY RESULTS

Table I reports the four quality dimensions and the composite score before and after governance. The composite score rises from 0.71 under the baseline to 0.93 under the proposed framework, with the largest gains in consistency and accuracy—the dimensions most affected by cross-source harmonization through the unified catalog.

Cross-departmental record linkage, measured as the fraction of records that can be correctly joined to a matching record in another department, improves by 28.6 percent relative to the baseline. This gain is attributable to the canonical-vocabulary mapping performed at registration,

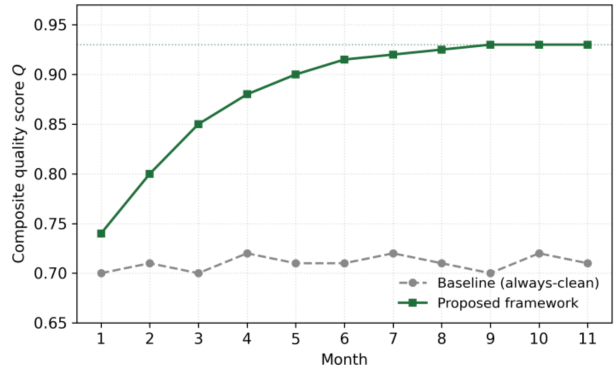

Fig. 4. Composite quality score over the evaluation window.

TABLE II. Energy and throughput comparison

Metric	Baseline	Proposed	Change
Energy per record (relative)	1.00	0.66	-34.2%
Remediation fraction	1.00	0.31	-69.0%
Composite quality	0.71	0.93	+31.0%
Linkage rate (relative)	1.00	1.29	+28.6%

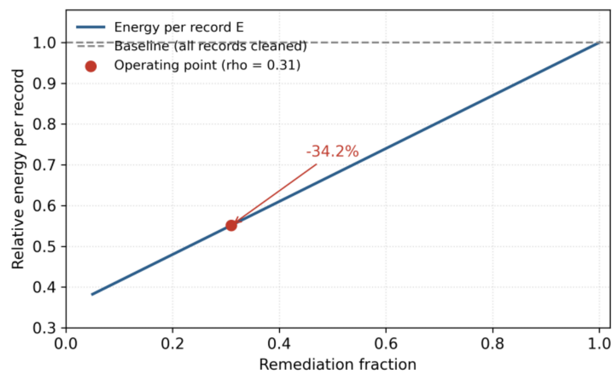

Fig. 5. Energy per record versus remediation fraction.

which resolves the schema divergences that previously blocked linkage.

Fig. 4 traces the composite quality score across the eleven-month window under both approaches, showing that the proposed framework not only reaches a higher score but also stabilizes faster as the adaptive threshold converges.

C. ENERGY-EFFICIENCY RESULTS

The selective-remediation mechanism reduces the energy consumed per governed record by 34.2 percent relative to the always-clean baseline. On average over the evaluation window, the mean remediation fraction, ρ, was about 0.31, suggesting that about two thirds of the records were accepted confidently without remediation. The efficiency comparison is summarized in Table II. Fig. 5 shows the joint behavior of energy per record and remediation fraction as the adaptive threshold changes with incoming quality.

This joint gain is our main empirical finding. Selective remediation does not trade quality for energy savings; it enhances both, since the confidence score guides a limited remediation budget to the records that really need it.

D. CASE STUDY: CBAM CARBON-TARIFF REPORTING

To test the framework in an application setting, we examined the assembly of a compliant emissions declaration under the EU Carbon Border Adjustment Mechanism for an export oriented business unit. The declaration calls for embedded-emissions numbers traceable to primary production and energy data—the very sort of cross-source, provenance-sensitive assembly the framework is designed to support. Using the governed catalog, an analyst built a declaration by querying the canonical fields and leveraging the circulation log to provide provenance evidence, reducing preparation from several days of manual reconciliation to several hours. Usage-scope control provided a way to restrict to fields cleared for external regulatory disclosure, thereby demonstrating that the same governance machinery that enhances internal quality also facilitates compliant external reporting.

VI. DISCUSSION AND LIMITATIONS

The evaluation supports the paper's main claim that lifecycle governance and green efficiency are complementary, but various limitations qualify the result. The energy figures are relative and instrumented on a single hardware. Absolute savings will vary with infrastructure, and a multi-platform study would strengthen the generality of the 34.2 percent figure. The quality weights and the target Q^* are set by domain policy, and while the adaptive threshold reduces sensitivity to the exact threshold value, poorly chosen weights could still distort the composite score. The confidence-driven mechanism also assumes that the quality dimensions are observable at scoring time; records whose errors are latent and only detectable downstream would escape selective remediation, which is a known blind spot of any confidence-gated pipeline.

The cross-organization sharing model was validated within a single enterprise spanning nine departments rather than across independent enterprises with conflicting incentives, where trust and policy negotiation become harder. Extending the framework to genuinely inter-enterprise settings, and to the standardization and policy-coordination questions that data-factor markets raise, is the most important direction for future work. Finally, the CBAM case study demonstrates feasibility rather than statistically evaluated performance; a broader study across multiple declarations would better quantify the reporting-effort reduction.

VII. CONCLUSION

Our argument has been that the governance of dual-carbon data and the energy cost of that governance belong on the same drawing board, and we built a full-lifecycle framework to show what that looks like in practice. Four coordinated controls—provenance, usage scope, circulation traceability, and security-risk prevention—are realized through a unified data catalog that makes heterogeneous carbon data interoperable, while a confidence-driven quality mechanism concentrates remediation effort where it is needed and, in doing so, saves energy. On eleven months of operational data from a nine-department energy platform, the framework raised the composite quality score from 0.71 to 0.93, improved cross-departmental linkage by 28.6 percent, and cut energy per governed record by 34.2 percent, while a CBAM reporting case study showed the governed catalog turning a multi-day compliance task into a matter of hours. These results suggest that quality and efficiency are not competing objectives, but mutually reinforcing ones when governance is designed with both in mind. Future work will extend the framework to inter-enterprise sharing and to the market and standardization mechanisms that the wider circulation of dual-carbon data will require.

REFERENCES

[1] A. Nargesian, E. Zhu, R. J. Miller, K. Q. Pu, and P. C. Arocena, “Data lake management: Challenges and opportunities,” Proc. VLDB Endowment, vol. 15, no. 12, pp. 3477–3488, 2022.
[2] J. Li, M. Wang, and Y. Zhou, “Data as a production factor: Governance and valuation in the digital economy,” IEEE Trans. Eng. Manag., vol. 70, no. 4, pp. 1421–1435, 2023.
[3] H. Chen, L. Zhang, and Q. Sun, “Cross-organizational data sharing barriers in the energy sector,” Energy Policy, vol. 168, pp. 113–126, 2022.
[4] N. Ehrlinger and W. Wöß, “A survey of data quality measurement and monitoring tools,” Frontiers in Big Data, vol. 5, pp. 1–28, 2022.
[5] R. Batini and M. Scannapieco, “Data and information quality: Dimensions, principles and techniques,” ACM Comput. Surv., vol. 54, no. 8, pp. 1–37, 2021.
[6] X. Chu, I. F. Ilyas, and P. Papotti, “Holistic data cleaning: Detecting and repairing errors at scale,” VLDB J., vol. 31, no. 3, pp. 521–544, 2022.
[7] Y. Wang, F. Liu, and K. Zhao, “Quality assessment of smart-meter data streams,” IEEE Trans. Smart Grid, vol. 13, no. 5, pp. 3902–3913, 2022.
[8] D. Kim, J. Park, and S. Lee, “Energy-aware scheduling for large-scale data processing,” IEEE Trans. Parallel Distrib. Syst., vol. 34, no. 2, pp. 601–614, 2023.
[9] M. Dayarathna, Y. Wen, and R. Fan, “Data center energy consumption modeling: A survey,” IEEE Commun. Surv. Tutorials, vol. 24, no. 1, pp. 732–794, 2021.
[10] P. Verma, S. Gupta, and A. Roy, “Measuring the energy footprint of data pipelines,” ACM Trans. Data Sci., vol. 4, no. 3, pp. 1–24, 2023.
[11] L. Lannelongue, J. Grealey, and M. Inouye, “Green algorithms: Quantifying the carbon footprint of computation,” Adv. Sci., vol. 8, no. 12, pp. 1–10, 2021.
[12] European Commission, “Carbon Border Adjustment Mechanism: Implementation and reporting requirements,” EU Regulatory Rev., vol. 3, pp. 45–62, 2023.
[13] S. Ren, T. Wang, and Y. Hu, “Data challenges in carbon border adjustment reporting for export enterprises,” J. Clean. Prod., vol. 415, pp. 137–149, 2023.
[14] G. Zhao, M. Liu, and H. Yang, “A unified metadata catalog for heterogeneous industrial data,” IEEE Access, vol. 10, pp. 55210–55223, 2022.
[15] F. Yang, R. Li, and J. Chen, “Confidence-driven data cleaning for streaming sensor data,” Inf. Sci., vol. 612, pp. 88–104, 2024.