Kern Public : Pages:1–10

Original Research

Challenges and Best Practices in Implementing Unified Data Architectures for Global Corporations

Nguyen Minh Quang[1] and Tran Duc Thien[2]

[1]Mekong University of Commerce, Business Department, Vo Van Kiet Street, Can Tho, Vietnam.
[2]Hanoi Institute of Business Studies, Business Department, Pham Hung Road, Hanoi, Vietnam.

Abstract
Global corporations are increasingly dependent on data to coordinate operations, meet regulatory obligations, and understand rapidly evolving markets. Over time, many organizations have accumulated fragmented data systems that reflect historical mergers, regional autonomy, and technology shifts such as the adoption of cloud platforms and software-as-a-service applications. These fragmented landscapes create obstacles for cross-border analytics, consistent reporting, and shared operational processes. Unified data architectures have emerged as one approach to addressing these challenges by bringing together disparate data stores, integration mechanisms, and governance practices into a more coherent structure. This paper examines practical challenges and best practices in implementing unified data architectures in global corporations, focusing on both technical and organizational dimensions. It considers issues such as aligning heterogeneous data models, integrating on-premise and cloud environments, designing governance structures that operate across jurisdictions, and managing change in complex stakeholder environments. The discussion also considers patterns in platform selection, data integration approaches, security and privacy considerations, and operating models for data teams. Rather than proposing a single prescriptive framework, the paper synthesizes recurring themes and trade-offs observed in practice, highlighting factors that tend to influence implementation outcomes. The goal is to provide a structured view of the main tensions and decision points that global organizations face when attempting to unify their data architectures across diverse regions, business units, regulatory contexts, and technology stacks.

1. Introduction

Large multinational corporations operate across many countries, business lines, and regulatory regimes, and their information landscapes often mirror this complexity [1]. Historical investments in enterprise resource planning systems, local customer relationship management platforms, and industry-specific solutions have resulted in heterogeneous data silos. Acquisitions introduce additional variation as newly acquired entities bring their own systems, data models, and integration patterns. Over time, the result is a patchwork of overlapping capabilities, duplicated data, and inconsistent semantics that hinder the ability to obtain a consolidated view of customers, products, operations, and risks. As expectations for real-time insight and coordinated decision making increase, these fragmented architectures become more difficult to sustain [2].

Unified data architectures are one response to this situation. They generally aim to provide a coherent environment in which data from multiple sources can be ingested, transformed, governed, and exposed for analytical and operational purposes. In practice, this may involve data warehouses, data lakes, lakehouse platforms, streaming infrastructure, master data management solutions, and integration tools connected through common governance, cataloging, and security services [3]. The emphasis is less on a single technology and more on a coordinated set of architectural principles, shared platform components, and common operating practices. For global organizations, the aspiration is to balance centralization of standards and shared services with local autonomy, regulatory compliance, and flexibility for business units to innovate.

Implementing such architectures is rarely straightforward. Technical questions about integration patterns, platform selection, and performance sit alongside organizational questions about funding models, responsibilities, and decision rights. Regional differences in regulation, such as data residency requirements and industry-specific privacy rules, further complicate design choices. The concept of a single global data platform can conflict with the realities of network latency, legal boundaries, and local operational dependencies. Consequently, many global corporations experience repeated cycles of centralization and decentralization in their data architectures, as strategies evolve in response to changing constraints and priorities.

This paper focuses on the challenges and best practices associated with implementing unified data architectures in global corporations [4]. The discussion does not advocate a particular product stack or organizational model. Instead, it aims to describe the main forms of complexity that organizations encounter, the trade-offs inherent in various architectural choices, and the practices that appear to mitigate common risks. The analysis distinguishes between conceptual design issues, such as defining architectural principles, and more concrete topics such as integration techniques, data governance mechanisms, and operating models for data teams.

Challenge	Description	Context
Model Misalignment	Differing schemas across regions	Legacy integrations
Latency Constraints	Slow cross-border transfers	Real-time needs
Regulatory Limits	Data residency restrictions	Regional compliance
Access Conflicts	Varying entitlement rules	Multi-team usage

Table 1. *Key Technical and Regulatory Challenges.*

Domain	Ownership	Scope	Notes
Customer	Global + Regional	Core identifiers	Common mapping
Product	Global	Master hierarchy	Variant extensions
Finance	Central	Consolidated metrics	Periodic updates
HR	Regional	Local attributes	Sensitive fields

Table 2. *Examples of Data Domains in Unified Architectures.*

Layer	Purpose	Typical Content
Raw	Traceability	Source extracts
Standardized	Harmonization	Aligned schemas
Curated	Consumption	Business-ready data
Operational	Immediate use	Workflow data

Table 3. *Common Layers in Unified Data Architectures.*

The following sections first outline conceptual foundations for unified data architectures in a global context, including typical target states and common misunderstandings. Subsequent sections examine organizational and governance challenges, technical and integration concerns, and a set of recurring design patterns and practices observed across different sectors. A further section considers implementation roadmaps, metrics, and approaches to iterative improvement. The paper concludes with reflections on the tension between unification and diversity in data architectures, and on the conditions under which unified approaches tend to be sustainable in large international organizations.

2. Conceptual Foundations of Unified Data Architectures

Unified data architectures are sometimes interpreted as attempts to consolidate all data into a single platform [2]. In practice, the notion of unification is more nuanced. For global corporations, unification typically refers to a consistent set of principles, shared services, and integration patterns that allow distributed data assets to be discovered, combined, and governed in a coherent manner. A unified architecture may encompass multiple physical platforms, including regional data stores, specialized analytical environments, and legacy systems that cannot be easily replaced. The unifying elements often include a common data catalog, shared identity and access management, standardized data quality practices, and agreed policies for data lifecycle management.

A conceptual distinction can be drawn between logical and physical unification. Logical unification focuses on providing a coherent view of data through standardized schemas, semantic models, and interfaces that hide underlying heterogeneity. Physical unification emphasizes consolidation of storage and processing in a smaller number of platforms or regions. Many global corporations pursue a hybrid approach, seeking logical unification where physical consolidation is constrained by regulatory, latency, or operational considerations [5]. Understanding this distinction is important, because stakeholders may have different expectations about what unification entails, and misaligned expectations can lead to dissatisfaction with implementation outcomes.

Another foundational consideration is the scope of the architecture. Some organizations define unified data architectures primarily in relation to analytical use cases such as reporting, business intelligence, and data science. Others extend the concept to include operational data integration, event-driven architectures, and data products embedded in transactional systems. For global corporations, the boundary between analytical and operational contexts is often blurred, as customer and operational data is reused for real-time decisioning, personalization, and risk monitoring. A clear articulation of scope helps avoid overextension of architectural initiatives and supports prioritization of use cases that benefit most from unification.

The concept of data domains also plays a role in conceptual foundation. Rather than treating data as a single monolithic asset, organizations increasingly group related data into domains such as customer, product, finance, supply chain, and human resources [6]. Each domain may have its own subject matter experts, governance structures, and data models, even as it participates in a shared architecture. In global corporations, domains may further subdivide along regional or business line boundaries, reflecting local regulatory requirements or market conditions. Unified architectures must accommodate this domain structure, supporting common standards while allowing domain-specific variation where it fulfills legitimate needs.

Time horizons are another important conceptual dimension. Unified data architectures cannot be fully realized within a short project timeline. They evolve through increments as existing systems are integrated, new platforms are introduced, and governance practices mature. A realistic view of time allows organizations to distinguish between target-state architectures and transitional patterns that coexist over several years. In global settings, staged approaches are often necessary because the readiness of regions and business units varies, and major system migrations may need to align with local contract cycles or regulatory milestones [7]. Conceptual models that explicitly incorporate transitional states can guide implementation decisions and reduce frustration when immediate consolidation is not feasible.

Finally, unified data architectures need to be framed in relation to business value without overstating their impact. While improved data integration can enable more reliable reporting and more flexible analytics, business outcomes also depend on factors such as process design, organizational capabilities, and change management. Conceptual narratives that present architecture as one contributor among multiple factors tend to support more balanced expectations. This framing is especially important in large global corporations where many stakeholders are involved, and where competing initiatives, budget constraints, and shifting strategic priorities can influence the trajectory of architectural programs.

3. Organizational and Governance Challenges in Global Contexts

Organizational and governance factors often have greater influence on the success of unified data architectures than purely technical decisions. In global corporations, the distribution of authority across headquarters, regional hubs, and local business units introduces multiple layers of decision making. These layers may have differing incentives and regulatory exposures [8]. Central teams may prioritize standardization, economies of scale, and global reporting, while local entities may focus on responsiveness to local customers, compliance with national regulations, and minimizing disruption to existing operations. Aligning these perspectives requires governance structures that make decision rights explicit and provide mechanisms for negotiation.

A common challenge is the allocation of funding and ownership for shared data platforms. When platforms are centrally funded, regions may perceive them as headquarters initiatives and question their relevance to local needs. When costs are allocated to regions, local leaders may prefer investments in directly visible business applications over contributions to shared infrastructure. Governance forums that include both central and regional representatives can help negotiate cost-sharing models, prioritize capabilities, and ensure that platform roadmaps reflect diverse needs. However, such forums need clear mandates and escalation paths, otherwise decision making can become slow and ambiguous.

Data ownership and stewardship also become complex in global corporations [9]. Data about customers, employees, suppliers, and products may be created and maintained in multiple regions and systems. Defining who is accountable for data quality, metadata, and access decisions can be challenging when legal responsibilities, contractual relationships, and operational realities are distributed. Stewardship models that assign responsibilities at the level of data domains and that recognize both global and local roles can provide clarity. For example, a global domain owner may define core attributes and standards, while regional stewards manage extensions and local compliance considerations. Such arrangements require supporting processes for resolving conflicts and maintaining shared definitions over time.

Regulatory and compliance requirements influence governance in significant ways. Data protection laws, financial reporting standards, and sector-specific regulations can vary by jurisdiction and sometimes impose conflicting obligations. A unified data architecture must therefore accommodate local constraints while enabling consolidated reporting where permitted [10]. Governance structures need to incorporate legal, compliance, and privacy experts who can interpret regulations and translate them into policies for data classification, retention, access, and cross-border transfers. The interpretation of regulations may evolve as authorities issue guidance or as courts clarify ambiguities, requiring governance processes that can adjust policies and technical controls without destabilizing the architecture.

Cultural differences across regions also shape governance practices. Expectations around hierarchy, decision making speed, and risk tolerance can vary within the same corporation. A governance mechanism that functions effectively in one region may be perceived as overly rigid or insufficiently formal in another. Communication styles and language differences can add further complexity to discussions about standards and data quality expectations. To address these challenges, many organizations adopt governance models that combine global principles with regional implementation flexibility. Global principles define non-negotiable requirements such as alignment with certain security controls or access management practices, while regions can select specific tools or workflows that align with their local context as long as they meet shared requirements [11].

The role of leadership is particularly important in sustaining governance over time. Unified data architectures often span multiple CIO organizations, business units, and functional areas such as finance and risk. When leaders at these levels provide consistent messaging about the purpose of the architecture, support compromise on contentious design decisions, and endorse shared metrics for data quality and adoption, governance forums are more likely to maintain momentum. Conversely, frequent leadership changes or conflicting priorities can lead to fragmented initiatives and erosion of standards. Sustained leadership attention helps keep governance aligned with broader corporate strategies and provides a basis for resolving trade-offs when resource constraints or competing demands arise.

4. Technical and Integration Challenges Across Heterogeneous Landscapes

From a technical perspective, unified data architectures in global corporations must address heterogeneity in systems, platforms, and integration methods that have accumulated over many years. Legacy mainframe applications, enterprise resource planning systems, specialized industry platforms, and newer cloud-based applications often coexist. Each may store data in different formats, implement distinct reference data, and expose varying integration interfaces [12]. Bringing these systems into a unified architecture requires a combination of batch extraction, real-time integration, and event streaming techniques, chosen with attention to latency requirements, system load characteristics, and operational risk.

Factor	Central View	Regional View	Tension
Cost	Shared efficiency	Local budgets	Allocation
Control	Uniform governance	Local autonomy	Negotiation
Security	Standard controls	Local rules	Variability
Performance	Global routing	Edge processing	Balancing

Table 4. *Organizational Tensions in Global Data Programs.*

Metric	Focus	Use
Quality Score	Data integrity	Trust signals
Adoption Count	Consumer activity	Usage tracking
Latency Measure	Pipeline timing	Operational insight
Coverage Ratio	Domain inclusion	Roadmap view

Table 5. *Illustrative Metrics in Unified Architectures.*

Pattern	Role	Benefit	Scope
Data Contracts	Stability	Reduced breakage	Cross-domain
Shared Catalog	Discovery	Easier access	Enterprise
Automated Policies	Enforcement	Consistent rules	Multi-region
Feedback Loops	Improvement	Iterative refinement	All teams

Table 6. *Recurring Architectural Patterns.*

Risk Type	Source	Mitigation
Concentration Risk	Shared services	Redundant zones
Compliance Drift	Changing rules	Periodic reviews
Pipeline Failure	Complex paths	Incremental testing
Access Misconfig	Multi-platform	Unified IAM

Table 7. *Representative Risks and Mitigations.*

Stage	Focus	Output	Notes
Pilot	Limited scope	Initial patterns	Regional cluster
Scale-Up	Broader reach	Added domains	Platform maturity
Stabilization	Quality tuning	Reliable flows	Monitoring
Expansion	New use cases	Extended coverage	Incremental growth

Table 8. *Typical Implementation Stages.*

One enduring technical challenge is the alignment of data models across systems. Even when datasets appear similar, such as customer records or product hierarchies, differences in key structures, attribute definitions, and code sets can impede straightforward integration. For example, one region may represent customers at the legal entity level, while another uses account-level identifiers or household groupings. Product codes may be reused in different contexts or extended locally in ways that are not immediately visible in central catalogs. Achieving a unified view requires mapping and reconciliation processes that are often complex and subject to change as source systems evolve. Data modeling practices that distinguish between global canonical models and local extensions can help manage this complexity, but they do not eliminate the need for ongoing alignment work.

Latency and performance considerations add further complications [13]. Some use cases, such as regulatory reporting, can rely on daily or even less frequent data refreshes. Others, such as fraud detection or personalized customer engagement, require near real-time data flows. Implementing a unified architecture that can support these varied latency profiles often involves combining batch-oriented data pipelines with streaming platforms. In global settings, network distance and cross-border bandwidth constraints can influence design decisions, especially when transmitting large volumes of data between regions. Caching, regional aggregation layers, and data minimization strategies are often employed to reduce network load while preserving the usefulness of the data for intended use cases.

Integration with cloud services introduces additional layers of complexity. Many global corporations adopt multiple cloud providers, either by design or as a result of acquisitions and regional mandates [14]. This can lead to a mix of native cloud data services, third-party platforms, and on-premise systems that must interoperate. Differences in identity management, storage semantics, and security controls across cloud providers require careful configuration to maintain a consistent governance posture. Decisions about whether to centralize data in a single cloud region, replicate it across multiple regions, or maintain a hybrid architecture with strong regionalization are influenced by costs, latency, regulatory requirements, and resilience considerations.

Data quality and observability are also central technical concerns. When data is sourced from many systems across regions, inconsistencies, missing values, and timing issues can undermine trust in unified datasets. Technical architectures need to embed data quality checks, lineage tracking, and monitoring to detect and address these issues. Data quality rules may need to be localized even when applied within a unified framework, because acceptable value ranges, valid code sets, and business rules can vary across markets. Observability extends beyond data quality to include monitoring of pipeline performance, error rates, and resource utilization across heterogeneous environments [15]. Consistent logging, metrics, and alerting frameworks are important for managing this complexity.

Security and privacy controls must be integrated into the architecture rather than bolted on as an afterthought. Access control models that align with corporate policies and local regulations need to function across platforms, identity providers, and integration mechanisms. Encryption at rest and in transit, key management, and tokenization may be required, especially for personal and sensitive data. In some jurisdictions, data localization rules limit cross-border data transfers or impose specific conditions on such transfers. Technical designs must support logical segmentation of data by region, flexible anonymization or pseudonymization mechanisms, and configuration options that allow regional security teams to enforce local policies while remaining consistent with global standards. Achieving this balance is technically demanding but necessary for a sustainable unified architecture.

5. Best Practices and Architectural Patterns for Global Unified Data Architectures

While each global corporation operates under specific constraints, a number of recurring patterns and practices can be observed in unified data architecture initiatives [16]. One widely adopted approach is to separate the architecture into shared core platforms and domain-specific data products. Shared core platforms provide common capabilities such as data ingestion frameworks, storage and compute resources, cataloging, lineage, and access management. Domain-specific data products are designed and owned by business or functional domains, which use the core capabilities to publish data in forms aligned with their use cases. This separation allows central teams to focus on providing robust foundational services, while domains address their particular semantic and analytical needs.

Another practice involves the adoption of layered architectures that distinguish between raw, standardized, and curated data zones. Raw zones store data in forms closely reflecting source systems, providing traceability and flexibility for reprocessing. Standardized zones apply transformations to align data with shared reference models, such as global customer or product schemas, while retaining sufficient detail for cross-domain integration. Curated zones contain datasets tailored to specific analytical or reporting needs, often with additional quality checks and business logic [17]. In global corporations, these layers might exist at both regional and central levels, with regional layers handling localization and central layers focusing on cross-regional aggregation and harmonization.

A further best practice is the explicit design of data contracts between data producers and consumers. Instead of treating data sources as passive repositories, organizations define interfaces and expectations around schema stability, data availability, and quality characteristics. In global contexts, contracts help coordinate across time zones, languages, and organizational boundaries. They can specify, for example, when certain datasets will be updated, which attributes are considered stable cores versus optional extensions, and how changes will be communicated. Data contracts do not remove all integration issues, but they provide a reference point for discussions and help reduce unexpected breakages when underlying systems evolve.

Standardization efforts tend to be more effective when they concentrate on a limited set of high-value areas rather than attempting to harmonize every attribute across all systems. For instance, shared identifiers for customers, products, and organizational units, along with standard code sets for key dimensions such as country and currency, often deliver significant integration benefits [18]. In contrast, forcing uniformity in every local attribute can lead to resistance and workarounds. Many global corporations therefore prioritize a core set of shared data elements that support cross-border processes and reporting, while allowing local systems to maintain additional attributes that are relevant to local operations or regulations.

Automated governance controls are another important feature of mature unified architectures. Instead of relying solely on manual approvals or documentation, organizations embed policies into platform capabilities. Examples include automated enforcement of access policies based on data classification, predefined retention rules that trigger archival or deletion processes, and continuous compliance checks that validate encryption configurations or cross-border data transfers. Automation is especially valuable in global settings where manual oversight would be difficult to sustain over numerous datasets, regions, and teams. However, automated controls need to be accompanied by clear exception-handling mechanisms and periodic review to ensure alignment with evolving regulations and business needs.

Finally, successful unified architectures often incorporate deliberate mechanisms for feedback and learning [19]. Data consumers, such as analysts, data scientists, and operational teams, can provide insights into gaps in the architecture, such as missing metadata, inconsistent definitions, or performance limitations. Structured feedback channels, such as user councils or regular review sessions, enable platform teams and governance bodies to adapt roadmaps and refine standards. Over time, these mechanisms support a more realistic match between architectural designs and actual usage patterns. In global corporations, feedback loops can also reveal regional differences in adoption and identify where additional support, training, or localized enhancements are needed to ensure that the unified architecture remains relevant across diverse contexts.

6. Implementation Roadmaps, Metrics, and Continuous Improvement

Translating unified data architecture concepts into operational reality requires structured implementation roadmaps. In global corporations, such roadmaps rarely proceed in a strictly linear manner, but they often share certain characteristics. A common pattern is to begin with a limited number of high-priority use cases that require cross-regional or cross-domain data integration, such as global regulatory reporting, consolidated financial analytics, or group-level risk monitoring. These early use cases provide a rationale for investing in shared platforms and governance structures while keeping the initial scope manageable [20]. As capabilities mature and trust in the architecture grows, additional use cases and domains can be onboarded.

Phased rollouts are often organized around regional clusters or business units. For example, an organization might initially focus on regions with relatively homogeneous systems or regulatory environments, using early deployments to refine integration patterns and governance practices. Lessons learned from these pilots can then inform deployments in more complex or tightly regulated regions. The sequencing of regions is influenced by factors such as business priority, technology readiness, availability of local champions, and the timing of ongoing system transformations. Clear articulation of phase boundaries and success criteria helps maintain momentum and avoid scope creep.

Metrics play a significant role in guiding implementation and assessing progress. Beyond technical metrics such as pipeline performance or storage utilization, organizations often track measures related to data quality, adoption, and process efficiency [21]. Examples include the proportion of critical data domains integrated into the unified architecture, the number of active data consumers, the reduction in manual reconciliation activities, and the time required to produce specific reports. These metrics need to be interpreted with caution because improvements can depend on factors beyond the architecture itself, such as process redesign or staffing levels. Nevertheless, consistent measurement provides a basis for informed discussion among stakeholders and can support adjustments to roadmaps and investment levels.

Change management is an integral component of implementation in global settings. Introducing unified data platforms and governance practices can affect daily work for many roles, including data engineers, analysts, application owners, and business users. Training programs, documentation, and communication campaigns are needed to explain new concepts, responsibilities, and tools. For regions accustomed to high degrees of autonomy, unified architectures may initially be perceived as constraints. Addressing such concerns requires transparent communication about decision rationales, opportunities for local input into standards, and acknowledgment of legitimate local requirements that need to be accommodated [22]. Change agents and local champions can help translate central messages into regional contexts and provide feedback to central teams.

Continuous improvement is necessary because both the technology landscape and business requirements evolve. New cloud services, analytical techniques, and regulatory developments can shift the feasibility and desirability of certain architectural patterns. Roadmaps that explicitly allocate capacity for experimentation and refactoring are better positioned to adapt over time. For example, organizations might periodically reexamine their choices regarding data storage formats, integration technologies, or access patterns in light of emerging practices and cost structures. In global corporations, continuous improvement also involves monitoring the distribution of capabilities and adoption levels across regions, identifying where additional support or platform enhancements may be needed.

Risk management accompanies all stages of implementation. Unified data architectures can introduce new forms of concentration risk, such as dependence on shared platforms, centralized identity providers, or single cloud regions [23]. They may also increase the potential impact of misconfigurations, because changes to shared components can affect many downstream consumers. Implementation roadmaps therefore need to incorporate resilience considerations, including backup and disaster recovery strategies, segregation of duties, and controlled change management processes. Pilot deployments, staged rollouts, and robust testing regimes help limit the impact of unforeseen issues. Over time, lessons from incidents and near misses can be fed back into platform design and operational practices, contributing to a more resilient unified data environment.

7. Conclusion

Global corporations face persistent challenges in managing fragmented data landscapes that reflect historical growth, regional autonomy, and evolving technology ecosystems. Unified data architectures offer a way to address some of these challenges by providing coherent frameworks for integrating, governing, and using data across regions and business units. The discussion in this paper has emphasized that unification does not necessarily imply full physical consolidation, but rather the establishment of shared principles, services, and patterns that allow diverse systems to operate more coherently. Distinctions between logical and physical unification, between central and local responsibilities, and between core standards and optional extensions provide useful lenses for navigating the complexity of global data environments [24].

The analysis has highlighted that organizational and governance dimensions are as important as technical choices. Allocating ownership and funding, defining data stewardship roles, interpreting regulatory requirements, and accommodating cultural differences all influence the design and sustainability of unified architectures. Technical work to align data models, manage latency, integrate on-premise and cloud environments, and embed security and privacy controls needs to proceed in parallel with these organizational efforts. Best practices such as layered data zones, domain-oriented data products, data contracts, automated governance controls, and structured feedback mechanisms appear to support more stable and adaptable architectures over time.

Implementation roadmaps in global corporations benefit from phased approaches that focus initially on targeted use cases and regions, accompanied by clear metrics, change management strategies, and risk management practices. Continuous improvement is necessary as technologies, regulations, and business priorities change. Rather than seeking a final end state, organizations can aim for architectures that are capable of evolving while maintaining coherence in key areas such as shared identifiers, governance rules, and security controls. The experiences of different sectors suggest that while unified data architectures cannot remove all complexity from global operations, they can provide a more predictable foundation for data-driven analysis and decision making when designed and managed with a balanced appreciation of both technical and organizational realities [25].

References

[1] L. Cheng, F. Liu, and D. D. Yao, "Enterprise data breach: causes, challenges, prevention, and future directions," *WIREs Data Mining and Knowledge Discovery*, vol. 7, 6 2017.

[2] K. S. HEGDE, "Building a global cross-regional data platform to centralize data for a global enterprise," *INTERNATIONAL JOURNAL*, vol. 13, no. 11, pp. 616–619, 2024.

[3] R. Shah, K. Mukherjee, A. Tyagi, S. K. Karnam, D. Joshi, S. P. Bhosale, and S. Mitra, "R2d2: Reducing redundancy and duplication in data lakes," *Proceedings of the ACM on Management of Data*, vol. 1, pp. 1–25, 12 2023.

[4] C. Baah, Y. Agyabeng-Mensah, E. Afum, and J. A. L. Armas, "Exploring corporate environmental ethics and green creativity as antecedents of green competitive advantage, sustainable production and financial performance: empirical evidence from manufacturing firms," *Benchmarking: An International Journal*, vol. 31, pp. 990–1008, 4 2023.

[5] Z.-M. Zadorozhnyy, V. Muravskyi, S. Yatsyshyn, and O. Shevchuk, "Accounting of wages with the use of biometrics to ensure cybersecurity of enterprises," *Financial and credit activity problems of theory and practice*, vol. 3, pp. 162–172, 6 2021.

[6] D. Allemang, J. Hendler, and F. Gandon, *Counting and sets in OWL.* ACM, 7 2020.

[7] Y. Qaseem, S. Samra, O. German, E. Gray, and M. K. Gill, "Self-reported awareness of retinopathy severity in diabetic patients," *Clinical ophthalmology (Auckland, N.Z.)*, vol. 14, pp. 2855–2863, 9 2020.

[8] D. A. George, J. F. Clewett, C. Birch, A. Wright, and W. Allen, “Development and accreditation of an applied climate education unit for sustainable land use in australia,” *Journal of Sustainable Agriculture*, vol. 29, pp. 87–108, 4 2007.

[9] V. Chavan and L. Penev, “The data paper: a mechanism to incentivize data publishing in biodiversity science,” *BMC bioinformatics*, vol. 12, pp. 1–12, 12 2011.

[10] E. a. S. Elamparuthi, “Implementation of a hybrid color image compression technique using principal component analysis and discrete tchebichef transform,” *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, pp. 5374–5387, 4 2021.

[11] J. Chen, W. Shi, Y. Liu, Q. Chen, P. Huang, and J. Li, “Big data analysis supports the research and pilot application of distribution network planning technology,” *IOP Conference Series: Earth and Environmental Science*, vol. 632, pp. 042047–, 1 2021.

[12] Q. Rong, *Analyse Different Angles of Editing Customer Information Data in Customer Relationship Management*, pp. 302–306. Springer Singapore, 8 2021.

[13] L. Briesemeister, W. Gustafson, G. Denker, A. Martin, K. Martiny, R. A. Moore, D. Pavlovic, and M. S. John, *HCI (29) - Policy Creation for Enterprise-Level Data Sharing.*, pp. 249–265. Germany: Springer International Publishing, 6 2019.

[14] V. Gole and S. Shiralkar, *Current Trends in Analytics and SAP’s Road Map*, pp. 1–17. Apress, 9 2020.

[15] null Lixinyu, “Research on management uncertainty solution based on environmental change by big data technology,” *Journal of Physics: Conference Series*, vol. 1345, pp. 022063–, 11 2019.

[16] . , . , and . , “A model of the impact of bankruptcy costs and specific risks on the long-term economic growth of high-tech enterprises,” *Organizer of Production*, pp. 56–65, 6 2023.

[17] Y.-H. Choi, M. echelle Grace Zaragoza, D. S. Han, and H.-K. Kim, “Big data and hadoop for productive social commerce: Trending in social media users and entrepreneurs,” *International Journal of Software Engineering and Its Applications*, vol. 10, pp. 419–430, 12 2016.

[18] D. Chorafas, *Enterprise Architecture and New Generation Information Systems - Enterprise Data Storage and Corporate Memory Facility*. CRC Press, 12 2001.

[19] Z. Worku, “The impact of over-regulation on small enterprises,” *Risk Governance and Control: Financial Markets and Institutions*, vol. 6, pp. 12–19, 8 2016.

[20] C. Forliano, P. D. Bernardi, A. Bertello, and V. Temperini, “Innovating business processes in public administrations: towards a systemic approach,” *Business Process Management Journal*, vol. 26, pp. 1203–1224, 5 2020.

[21] K. D. Brunisholz, E. Joy, M. Hashibe, L. H. Gren, L. A. Savitz, S. Hamilton, W. Cannon, and J. Kim, “Incidental risk of type 2 diabetes mellitus among patients with confirmed and unconfirmed prediabetes.,” *PloS one*, vol. 11, pp. e0157729–, 7 2016.

[22] S. Jenkins, R. Bhanugopan, and P. Lockhart, “A framework for optimizing work–life balance practices in australia: Perceived options for employee support,” *Journal of Employment Counseling*, vol. 53, pp. 112–129, 8 2016.

[23] N. Scibelli, A. Mangano, K. Raynor, S. Wilson, and P. Singh, “A retrospective review of upper gastrointestinal bleed outcomes during hospital admission while on oral anticoagulation.,” *Cureus*, vol. 13, pp. e15061–, 5 2021.

[24] K. J. O’Leary, V. K. Devisetty, A. R. Patel, D. Malkenson, P. Sama, W. K. Thompson, M. P. Landler, C. Barnard, and M. V. Williams, “Comparison of traditional trigger tool to data warehouse based screening for identifying hospital adverse events,” *BMJ quality & safety*, vol. 22, pp. 130–138, 10 2012.

[25] F. Wang and J. Aviles, “Enhancing operational efficiency: Integrating machine learning predictive capabilities in business intellgence for informed decision-making,” *Frontiers in Business, Economics and Management*, vol. 9, pp. 282–286, 5 2023.