International Journal of Scientific Interdisciplinary Research
Vol 6, No 01, March 2025
Page 97-136
Doi: 10.63125/527rnx08

DATA PRIVACY IN BUSINESS INTELLIGENCE SYSTEMS: ENSURING COMPLIANCE IN HRIS AND ENTERPRISE PLATFORMS

Md Mostafizur Rahman[1]; Omar Muhammad Faruk[2]; Sheratun Noor Jyoti[3];

[1]. Master of Science in Management Information Systems, Lamar University, Texas, USA; Email: mrahman70@lamar.edu
[2]. MBA in International Business; Ajou University, Suwon, South Korea; Email: mohammad24omar@gmail.com
[3]. MA in Information Technology Management, Webster University-Saint Louis, MO, USA; Email: sheratunnoor@gmail.com

Abstract
This systematic literature review examines how data privacy can be engineered and governed as an intrinsic property of business intelligence (BI) programs that consume Human Resource Information Systems (HRIS) and wider enterprise platforms. Following a registered PRISMA protocol, we searched multidisciplinary databases, screened records in duplicate, appraised quality with design-appropriate tools, and synthesized heterogeneous evidence narratively and thematically. In total, 115 peer-reviewed articles and standards met the inclusion criteria. The synthesis maps regulatory obligations such as purpose limitation, minimization, storage limitation, transparency, rights handling, accountability, and integrity-confidentiality to concrete controls across the analytics lifecycle. We find that privacy outcomes improve when governance is encoded in code via policy-as-code approvals, slim ingestion schemas, retention rules, and purpose tags; least-privilege access is most durable when role-based baselines are refined with attribute-based context and enforced as row- and column-level security at the semantic layer; protection works best as a choreography that assigns encryption, tokenization, pseudonymization, and masking to precise pipeline stages; and verification hinges on engineered lineage, runtime observability, and tamper-evident audit logs that make "what happened to whose data and why" answerable under audit. We also surface operational patterns for DSAR fulfillment, retention and erasure in warehouses and lakehouses, machine unlearning for models, and cross-border and vendor risk mitigations using customer-managed keys, regionalization, and purpose-aware authorization. Contributions include a regulation-to-engineering crosswalk, a reference architecture, a maturity model, DPIA prompt templates tailored to analytics, and a compact assurance metric set for continuous monitoring. Collectively, the review translates legal principles into testable, auditable design choices that enable compliant, trustworthy HRIS-to-BI analytics at scale.

Keywords
Data privacy; Business intelligence; HRIS; GDPR compliance; Data governance; RBAC/ABAC; Row-level Security;

Citation:
Rahman, M. M., Faruk, O. M., & Jyoti, S. N. (2025). Data privacy in business intelligence systems: Ensuring compliance in HRIS and enterprise platforms. *International Journal of Scientific Interdisciplinary Research,* 6(1), 97–136. https://doi.org/10.63125/527rnx08

Received:
January 20, 2025

Revised:
February 18, 2025

Accepted:
March 17, 2025

Published:
April 05, 2025

INTRODUCTION

Business intelligence (BI) refers to the managerial processes, data infrastructures, and analytical applications that transform raw organizational data into decision-ready insights across functions and geographies (Chen et al., 2012). Human Resource Information Systems (HRIS) denote integrated platforms that record, process, and analyze employee-related information for operational and strategic human resource management, including payroll, timekeeping, performance, and talent analytics (Marler & Boudreau, 2017; Tursunbayeva et al., 2021). Within this landscape, privacy is the set of constraints governing the collection, use, disclosure, and retention of identifiable information in ways that respect contextual norms and individual expectations (Nissenbaum, 2004). When HRIS becomes a feeder to enterprise data warehouses and BI self-service tools, the same fields that support HR compliance and payroll accuracy can later appear in cross-functional dashboards or models, which raises new compliance obligations and audit needs in multinational settings. International significance emerges because global employers routinely synchronize HRIS, ERP, CRM, and collaboration telemetry across borders, where lawful bases, purpose limitation, data minimization, and storage limitation must be demonstrated with evidence in audits and DPIAs (Abraham et al., 2019; Cavoukian, 2010). Empirical and conceptual work on workplace monitoring shows that analytics over fine-grained activity logs heightens transparency and proportionality requirements, especially for people analytics (Ball, 2010, 2021; Marin & Sorgner, 2021). In healthcare and public sector deployments, HRIS implementations are intertwined with sectoral confidentiality regimes while still expected to enable workforce optimization and reporting. These definitional anchors frame BI/HRIS privacy as a socio-technical property requiring alignment of legal principles, engineering controls, and governance controls along the end-to-end data lifecycle.

Across jurisdictions, regulation shapes analytics design, documentation, and accountability. The GDPR establishes general obligations for lawfulness, fairness, and accountability, while acknowledging Member-State rules for employment-context processing and encouraging DPIAs where analytics present high risk. Organizational privacy engineering translates these obligations into role-segregated schemas, legal-basis registries, data minimization at ingestion, and retention enforcement within warehouses and BI workspaces (Chen et al., 2009). Access control research supplies models for constraining visibility: role-based access control operationalizes least privilege through roles and permissions, whereas attribute-based access control expresses policy in terms of user, resource, and environment attributes (Hu et al., 2015; Sandhu et al., 1996). Cloud adoption further foregrounds shared-responsibility alignment between controllers and processors and underscores the need for encryption, auditing, and secure regions in cross-border pipelines (Zissis & Lekkas, 2012). Guidance on de-identification clarifies terminology and technique families relevant to both structured and semi-structured releases used for analytics or benchmarking (Garfinkel, 2015). In HRIS-to-BI flows, these mechanisms converge with documentation practices policy-as-code, lineage, and approvals that make lawful bases and purpose constraints auditable at scale (Simmhan et al., 2005). Together, this body of work situates compliance not as a separate overlay but as a property synthesized from architectural choices, procedural controls, and continuous evidence-generation supporting oversight.

A large stream of scholarship examines identifiability and disclosure risk that can persist in high-dimensional enterprise data. Classical privacy models for data release k-anonymity, ℓ-diversity, and t-closeness seek to reduce the probability of linkage and attribute disclosure by generalizing, suppressing, or balancing sensitive attributes (Li et al., 2007; Machanavajjhala et al., 2007; Sweeney, 2002). Re-identification studies demonstrate that sparse behavioral or transactional records can uniquely single out individuals even when direct identifiers are removed, a finding replicated in mobility patterns and credit-card metadata. Methodological advances propose calibrating output noise to the sensitivity of queries and addressing local contributions and composition effects, providing defensible guarantees about incremental disclosure from repeated analytics . Surveys and monographs catalog threat models, utility metrics, and algorithmic techniques that practitioners can use to evaluate trade-offs in publishing or sharing datasets used for BI or external collaborations. For HRIS-aligned BI, these insights matter because person-level attributes job codes, tenure, pay bands, location, schedule patterns may combine with operational logs and auxiliary data to increase singling-out risk unless designs include cohort thresholds, aggregation, and provable privacy where appropriate (Duchi et al., 2013).

International Journal of Scientific Interdisciplinary Research
Vol 6, No 01, March 2025
Page No: 97-136
Doi: 10.63125/527rnx08

Figure 1: Systematic Framework for Privacy-Engineered HRIS-to-BI Data Flows

Technical safeguards within enterprise systems operate across multiple layers of the stack, weaving together a tapestry of protective mechanisms that ensure data integrity, confidentiality, and compliance. At the foundational database and semantic layers, authorization views and query rewriting techniques serve as powerful instruments for enforcing fine-grained access policies, allowing analysts and business intelligence (BI) consumers to view only those attributes and records explicitly permitted by policy (Ara et al., 2022; Rizvi et al., 2004). These controls are strengthened by access models such as Role-Based Access Control (RBAC), which organizes access rights according to organizational roles, and Attribute-Based Access Control (ABAC), which introduces a richer paradigm by binding access and decryption privileges to specific attributes, contextual conditions, and organizational rules, often applied consistently across microservices and distributed data products (Jahid, 2022; Shokri et al., 2011). Alongside these, encryption strategies further segment exposure, with attribute-based encryption and related cryptographic methods enabling policy-driven decryption for datasets exchanged between internal services or shared externally with trusted partners (Akter & Ahad, 2022; Zhang et al., 2020). Beyond traditional cryptography, privacy-preserving computation strengthens collaborative analytics: secure multiparty computation makes it possible for multiple stakeholders to conduct joint analyses without revealing their private inputs, while fully homomorphic encryption, despite its performance costs, allows computation directly over encrypted data (Arifur & Noor, 2022; Yao, 1982). Complementing these methods, differential privacy introduces statistical safeguards by injecting controlled noise into outputs, ensuring that aggregate queries can be answered with bounded disclosure risks, a principle that has already been embodied in large-scale industrial systems for telemetry and platform-level query engines (Erlingsson et al., 2014; McSherry, 2009; Hasan & Uddin, 2022). In human resource information systems (HRIS) integrated with BI pipelines, layered deployments such as RBAC and ABAC at the access level, encryption both at rest and in transit, policy-aware views at the modeling layer, and differential privacy or thresholding mechanisms at the aggregation stage collectively operationalize the principles of purpose limitation

and data minimization, translating them into executable, auditable controls that safeguard sensitive organizational information(Rahaman, 2022).

Governance and observability make these measures accountable. Data governance research articulates decision rights, stewardship, standards, quality, and oversight mechanisms that enable repeatable, auditable analytics processes across business units (Abraham et al., 2019; Rahaman & Ashraf, 2022). Provenance and lineage techniques document transformations from HRIS source tables to modeled facts and dimensional attributes, thereby supporting DSAR responses, retention enforcement, and DPIA documentation for analytics projects (Buneman et al., 2001; Islam, 2022). BI scholarship links governance to performance and adoption, emphasizing the need for metadata, cataloging, and standardized definitions so that privacy constraints and quality rules remain coherent across dashboards and data products (Chen et al., 2009; Hasan et al., 2022). De-identification guidance in enterprise contexts highlights the role of risk assessments, expert determination methods, and release management practices that balance analytic utility with acceptable disclosure risk (Garfinkel, 2015; Redwanul & Zafor, 2022). In practical HRIS scenarios, these governance controls determine which attributes are minimized at ingestion, which are masked or tokenized during modeling, which aggregates require thresholding, and which exports need additional contractual and technical controls during sharing.

People analytics sits at the intersection of organizational behavior, information systems, and law. Reviews and conceptual analyses map how algorithmic scoring and monitoring can affect autonomy, dignity, and fairness, thereby intensifying the importance of transparency, justification, and access discipline when designing HR dashboards and predictive models (Diakopoulos & Friedler, 2021; Garfinkel, 2015). Workplace surveillance work documents the breadth of monitoring from keystrokes to geolocation and communications and the corresponding requirements for proportionality, necessity, and clarity in organizational governance (Ball, 2010, 2021; Rezaul & Mesbaul, 2022). Contextual integrity provides an interpretive lens for understanding why the same attribute may be legitimate within an HR investigation but misaligned in broader operational analytics if transmission principles, actors, and purposes diverge (Hossen & Atiqur, 2022; Nissenbaum, 2004). In BI environments, these concerns materialize in access reviews, limited visibility for small cohorts, careful selection of outcome variables, and restrictions on free-text or communications analytics where sensitive inferences are likely. The literature thus equips practitioners to construct HRIS-to-BI programs that honor contextual norms through targeted governance levers role scoping, attribute filters, query auditing, and defensible aggregation while keeping analytic objectives anchored to lawful bases and explicit purposes (Kairouz et al., 2019; Solove, 2006; Warner, 1965).

Sectoral and international perspectives demonstrate how diverse safeguards, governance models, and technical strategies converge into coherent practice within human resource information systems (HRIS) and business intelligence (BI) environments. In healthcare, for instance, HRIS adoption cannot be isolated from the ethical and operational constraints of clinical governance; privacy and security strategies must be deliberately integrated into these frameworks while simultaneously enabling workforce analytics and regulatory reporting that support system-wide efficiency (Källander et al., 2017; Tawfiqul et al., 2022). Within Europe, employment data scholarship emphasizes the interplay between the General Data Protection Regulation (GDPR) and the varied labor regimes of member states, illustrating why multinational HRIS and BI infrastructures demand configurable policy engines and local approval workflows before analytics use cases can be deployed, ensuring lawful processing and worker protection (Hert & Papakonstantinou, 2022; Hasan, 2022). Complementing these legal imperatives, technical research on cloud security and cross-border data transfer provides the foundation for auditable controller–processor arrangements, where encryption, geographic region selection, and logging practices enable compliance with jurisdictional expectations (Tarek, 2022; Zissis & Lekkas, 2012). Data Protection Impact Assessment (DPIA) methodologies further refine this practice by prescribing systematic steps such as mapping data flows, conducting necessity analyses, and selecting safeguards that can be embedded directly into analytics delivery pipelines through automated lineage tracing, access governance, and retention policies (Kamrul &Omar, 2022; Simmhan et al., 2005; Tawalbeh et al., 2022). To address the risks of data reuse in reporting, privacy-preserving analytics introduces techniques such as differential privacy, thresholding, and budget-aware composition, thereby ensuring that sensitive HR statistics can be disclosed with quantifiable limits on disclosure risk (Dwork, 2006; Dwork & Roth, 2014; Kamrul

& Tarek, 2022). Orchestrating these diverse elements requires governance frameworks that align the roles of HR, Legal, Security, and Analytics stakeholders, establishing clear lines of accountability through documented approvals, audits, and outcome records, all of which are indispensable when BI platforms repurpose HRIS data to generate broader enterprise insights while still upholding compliance and trust (Garfinkel, 2015; Mubashir & Abdul, 2022).

The objective of this literature-based study is to produce a rigorous, practice-ready synthesis that clarifies how data privacy can be systematically engineered and governed within business intelligence environments that consume Human Resource Information Systems and other enterprise platforms. Specifically, the study aims to: define a precise vocabulary for HRIS-to-BI data flows, actors, and artifacts to remove ambiguity in technical and policy discussions; map the full set of privacy obligations that apply to analytics over workforce and enterprise data, including purpose limitation, minimization, storage limitation, accountability, access rights, erasure, portability, security safeguards, and transparency; catalogue the technical and organizational controls that appear across the peer-reviewed record such as role- and attribute-based access control, row- and column-level security, masking, tokenization, pseudonymization, encryption at rest and in transit, de-identification and disclosure control, differential privacy, secure computation techniques, and auditable lineage; evaluate the reported effectiveness, feasibility, and limitations of these controls in BI pipelines with attention to cohort size, join paths, auxiliary information, export practices, and self-service tooling; surface governance patterns that anchor privacy in day-to-day analytics work, including stewardship, data cataloging, policy-as-code, change control, approvals, and monitoring; construct a compliance-aware reference architecture that places control points along ingestion, modeling, storage, semantic, analysis, sharing, and deletion stages; develop a maturity model and a control-to-obligation matrix that practitioners can use to prioritize safeguards for HR dashboards, reports, data marts, experiments, and AI-assisted insights; examine operational problem areas data subject requests in aggregated datasets, retention and erasure in time-travel or snapshot storage, small-cohort reporting thresholds, export governance, and cross-border transfers with vendors and subprocessors; compare access-control strategies and privacy-preserving analytics choices against realistic BI use cases to articulate selection criteria; and identify evidence gaps, measurement needs, and reproducible evaluation methods that future investigations can adopt. The study's deliverables include a transparent review protocol, a thematic synthesis aligned to these objectives, a set of practitioner checklists, DPIA prompt templates tailored to analytics projects, and a minimal set of metrics for continuous assurance (for example, coverage of lineage, coverage of row-level security, coverage of retention rules, and mean time to complete erasure across derived assets.

LITERATURE REVIEW

The literature on data privacy within business intelligence (BI) environments that draw from Human Resource Information Systems (HRIS) and broader enterprise platforms spans interlocking streams that together define what "compliance-aware analytics" means in practice. Foundational work establishes the concepts that anchor the field privacy principles, contextual norms, data governance, and security models and explains how these shape the movement of data from HRIS sources into warehouses, semantic layers, and self-service dashboards. A second body of research details the regulatory and organizational context in which analytics unfolds, showing how lawful basis, purpose limitation, data minimization, storage limitation, accountability, and transparency must be evidenced through processes such as data protection impact assessments, stewardship roles, and audit trails. Technical studies contribute mechanisms that operate at different layers of the pipeline: access control (role- and attribute-based), row- and column-level security, masking and tokenization, pseudonymization, encryption at rest and in transit, and privacy-preserving analytics approaches such as differential privacy and secure computation; these are frequently coupled with lineage, observability, and logging that make controls inspectable. Complementing these safeguards is a robust stream on identifiability and disclosure risk in high-dimensional organizational data, which clarifies where re-identification can occur in aggregates or joins and motivates thresholds, cohort rules, and formally private query answering for sensitive metrics. The enterprise implementation literature contributes patterns and anti-patterns around modeling, semantic governance, export practices, and collaboration, highlighting the role of policy-as-code, change control, and approvals in day-to-day analytics work. Sector-specific analyses (for example, health systems and public administration) bring granularity to workforce-data use cases and demonstrate

how domain norms and contractual obligations interact with platform capabilities. Finally, evaluation-oriented studies propose criteria and metrics coverage of lineage, coverage of access controls, adherence to retention schedules, and time to fulfill deletion or access requests that allow organizations to measure whether privacy is genuinely embedded across ingestion, modeling, analysis, sharing, and deletion. Together, these streams provide a cumulative, multidisciplinary evidence base for structuring the review, benchmarking techniques and controls, and articulating a reference architecture and maturity model for privacy by design in HRIS-to-BI pipelines.

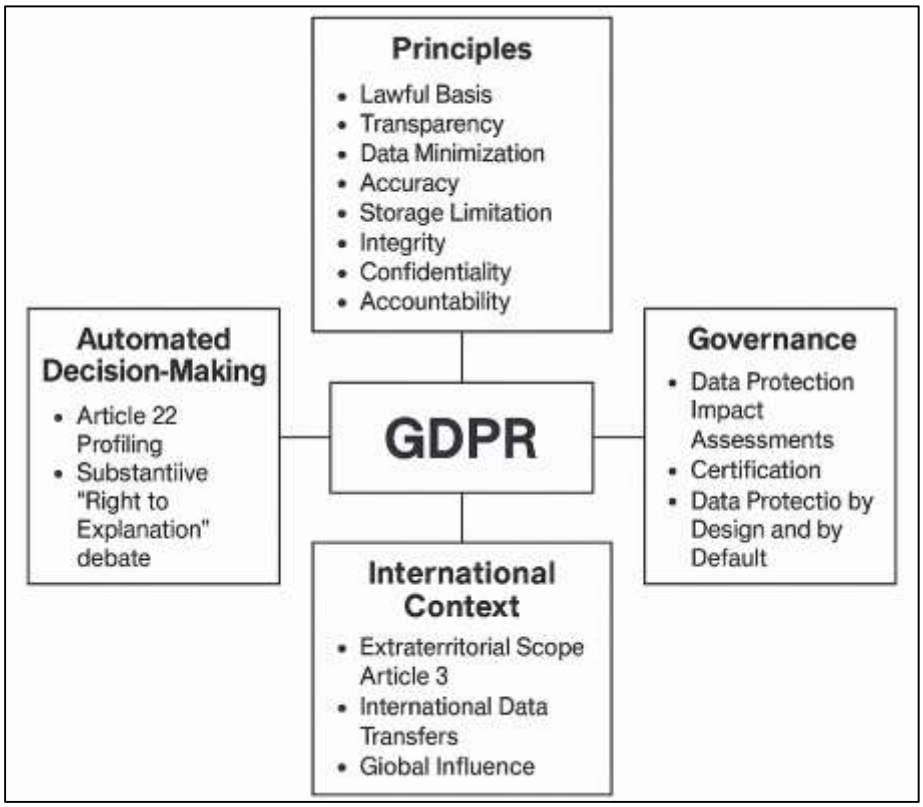
Business Intelligence (BI) over HRIS and Enterprise Data

Across business intelligence (BI) programs that integrate data from human resource information systems (HRIS) and wider enterprise platforms, the European Union's General Data Protection Regulation (GDPR) emerges as the dominant regulatory foundation, setting global standards for lawful basis, transparency, data minimization, accuracy, storage limitation, integrity, confidentiality, and accountability. These core principles apply directly to analytics that profile employees or applicants, regulate enrichment with third-party data, and restrict the downstream repurposing of operational datasets into dashboards, predictive algorithms, and performance models. A pivotal aspect is the intersection of the GDPR's fairness and transparency requirements with automated decision-making, which compels organizations to examine carefully when BI outputs may "significantly affect" individuals, thereby activating Article 22 safeguards and human-in-the-loop obligations. Scholarly debate continues over whether the GDPR enshrines a substantive "right to explanation," yet the practical consensus emphasizes providing meaningful information about the logic, impacts, and safeguards behind analytics while ensuring contestation pathways within HR workflows. Complementary regulatory guidance on profiling further illuminates how activities such as risk-tiering, performance scoring, and attrition prediction may qualify as regulated profiling, thereby mandating enhanced transparency, explicit accountability, and supervisory oversight (Borgesius, 2016; Lachaud, 2018; Muhammad & Kamrul, 2022). Adding to this, foundational doctrine around "singling out" warns system designers that even de-identified or pseudonymized records used for BI can fall under data protection scrutiny if individuals remain reasonably identifiable through linkage or inference, thereby underscoring the fragility of anonymity in analytics contexts. Collectively, these interpretive streams reframe HRIS-to-BI pipelines not as neutral technological processes but as regulated processing activities, where purpose limitation, role allocation between controller and processor, proportionality of data use, and explicit safeguards against discriminatory or opaque outcomes are not compliance afterthoughts but fundamental design constraints that shape metrics, model development, and data mart construction within enterprise BI architectures (De Hert & Czerniawski, 2016; Reduanul & Shoeb, 2022).

Beyond abstract principles, the GDPR gives governance concrete form through obligations that directly influence the architecture and execution of business intelligence (BI) programs. Central among these is the requirement for data protection impact assessments (DPIAs), which move beyond symbolic policy language to serve as structured, early-stage risk evaluations for projects likely to pose elevated risks such as large-scale monitoring, workforce profiling, or the handling of special-category HR data. These assessments oblige organizations to identify potential threats, evaluate their likelihood and severity, propose mitigation strategies, and, when residual risk cannot be sufficiently reduced, consult supervisory authorities before deployment. Empirical implementation studies show that organizations interpret these duties pragmatically by documenting lawful basis selections, designing explicit retention schedules, and clarifying role divisions across HR, People Analytics, Security, and IT, thereby embedding compliance into cross-functional governance practices. Alongside DPIAs, two governance levers have gained prominence in BI contexts: certification and the principle of data protection by design and by default (Goddard, 2017; Lachaud, 2020; Sabuj Kumar & Zobayer, 2022). Certification has been conceptualized as a form of "monitored self-regulation," extending beyond superficial conformity badges to embed oversight, accountability, and market discipline within compliance frameworks. This framing allows BI leaders to assess whether sector-specific schemes often aligned with ISO/IEC management standards can credibly assure controller and processor(Sadia & Shaiful, 2022).

International Journal of Scientific Interdisciplinary Research
Vol 6, No 01, March 2025
Page No: 97-136
Doi: 10.63125/527rnx08

Figure 2: BI over HRIS and Enterprise Data under the GDPR



performance while reinforcing vendor due diligence across the data supply chain. At the same time, the mandate of privacy by design and by default reorients BI engineering toward architectures that prioritize privacy as a default state, embedding measures such as attribute minimization in data models, privacy budgeting mechanisms in analytics workflows, carefully scoped role-based access, and privacy-conscious derivations of metrics. Scholarly contributions in this area bridge legal imperatives with technical governance, translating regulatory mandates into actionable design strategies and patterns that BI teams can instantiate in extract–transform–load (ETL) pipelines, semantic layers, and model-serving infrastructures (Hoofnagle et al., 2019; Sazzad & Islam, 2022; Wachter et al., 2017).

The extraterritorial scope and international data transfer regime of the GDPR impose further conditions on the design and governance of BI deployments within global enterprises, particularly when these systems draw on HRIS data. Article 3 extends the regulation's obligations to non-EU entities that either provide services to, or monitor the behavior of, EU data subjects, which means that even HRIS or BI platforms physically hosted or administered outside the European Union may still fall within the GDPR's jurisdictional reach. Legal scholarship on Article 3 highlights that determining applicability requires more than a simple location-based scoping of analytics projects, instead demanding careful analysis of targeting, monitoring, and processing contexts that link back to EU data subjects (Jasmontaite et al., 2018; Veale & Edwards, 2018). Cross-border data transfer jurisprudence since the Schrems cases has further recalibrated reliance on mechanisms such as Standard Contractual Clauses, while mandating risk assessments that compel organizations to revisit where HR data warehouses, BI caches, and event log pipelines are situated, and to adopt supplementary safeguards contractual, organizational, and technical that preserve equivalent protection for EU data once it leaves the Union (Noor & Momena, 2022). In parallel, scholarly and regulatory commentary positions the GDPR as a global benchmark whose influence extends into non-EU jurisdictions and corporate practices, harmonizing baseline expectations for transparency, accountability, and rights such as access and portability across multinational BI programs (Kuner, 2017; Akter & Razzak, 2022). For BI teams, this extraterritorial force translates into engineering obligations: building data architectures that are portability-ready to efficiently fulfill employee

access or portability requests, clearly mapping controller and processor responsibilities across layered SaaS and vendor ecosystems and anticipating supervisory scrutiny of profiling and automated decision-making outputs even when they are framed as internal analytics (Adar & Md, 2023). Ultimately, the intertwining of scope, transfer restrictions, and the GDPR's standard-setting role demonstrates that legal compliance is inseparable from the design of BI data architectures, vendor strategies, and model lifecycle management in HR-centered environments.

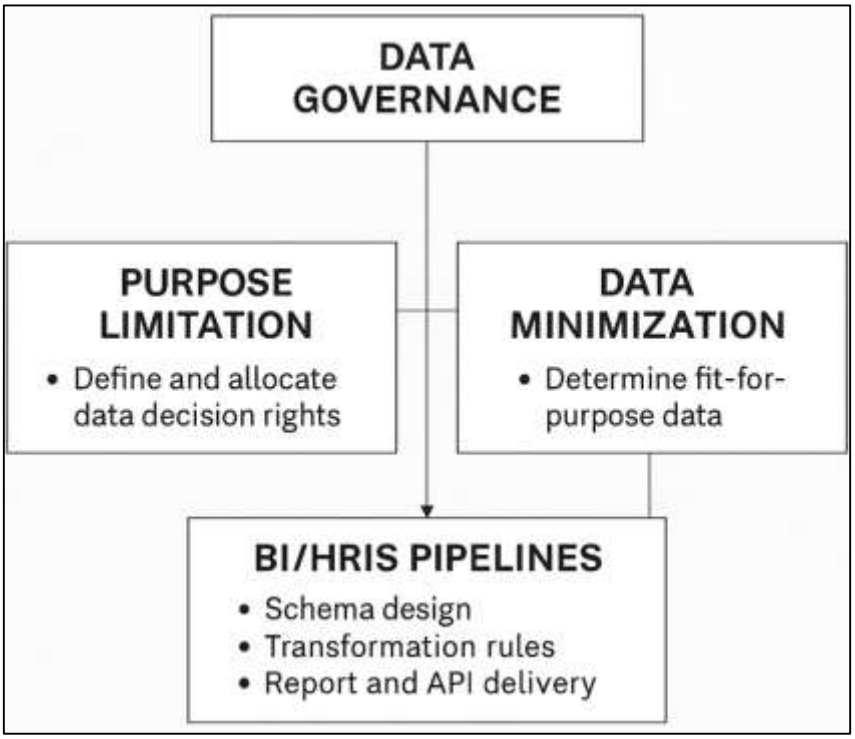
Data governance and data minimization in BI/HRIS pipelines

Robust data governance functions as the organizational cornerstone for embedding core privacy principles most notably purpose limitation and data minimization within interconnected business intelligence (BI) and HRIS ecosystems. At its heart, governance requires the enterprise to clearly define what types of data decisions must be made and to allocate rights and duties across the entire lifecycle, spanning ingestion, storage, transformation, and reporting. A well-regarded governance blueprint articulates these decisions across domains such as data principles, data quality, access management, lifecycle stewardship, and system architecture, each tied explicitly to loci of accountability, a framing especially suitable for cross-functional BI programs that integrate HR, finance, and operations data into unified reporting and analytics layers (Qibria & Hossen, 2023; Khatri & Brown, 2010). Complementing this, contingency-based thinking demonstrates that governance cannot be reduced to a rigid template; rather, it must be tailored to organizational realities, such as structural complexity, degree of process harmonization, and external regulatory demands pressures that are often heightened in HRIS landscapes where sensitive personal data routinely traverse multiple platforms (Istiaque et al., 2023; Weber et al., 2009). From a practitioner's perspective, distilled "catchword" style guidance proves valuable, as it translates abstract governance principles into assignable roles, RACI-style responsibilities, and escalation pathways that can be applied consistently across analytics projects, ensuring that BI initiatives remain aligned with purpose limitation even under pressure to expand scope (Akter, 2023; Otto, 2011). In more recent scholarship, system-level governance frameworks designed for trustworthy AI extend these traditions by emphasizing stewardship not only over data, but also over processes and algorithms, advocating controlled openness and proportionate, risk-based controls that balance innovation with accountability. Such approaches are particularly pertinent in modern BI architectures where HR workflows are increasingly enriched with machine learning, making governance a multi-layered discipline that safeguards privacy while enabling analytic value (Janssen et al., 2020; Hasan et al., 2023).

Operationalizing governance for data minimization begins by interrogating what constitutes "fit-for-purpose" data within BI and HRIS contexts, recognizing that the legitimacy of processing cannot rest on accuracy alone. Seminal work on data quality highlights that data consumers such as BI analysts and HR decision-makers evaluate utility through multiple dimensions, including relevance, timeliness, completeness, and interpretability, thereby underscoring that accuracy is only one facet of usable information (Masud et al., 2023; Wang & Strong, 1996). Framing these dimensions as preconditions for lawful processing enables organizations to argue convincingly for collecting less rather than more, since any field that does not contribute incremental utility for an approved analytic purpose should by default be excluded or aggregated. Provenance scholarship extends this logic by adding the crucial dimension of "explainability of origins," emphasizing that effective governance requires the ability to trace where data originated, how it was transformed, and why it appears in a given output. This why/how/where provenance not only reinforces accountability but also provides the audit trails BI and HRIS teams need to demonstrate alignment between input data, transformations, and declared purposes (Cheney et al., 2009; Sultan et al., 2023). Building on these foundations, governance activity maps presented in the literature serve as practical blueprints, spanning domains such as policy definition, stewardship responsibilities, quality controls, metadata management, access control design, and monitoring frameworks. These structured maps provide BI architects and HRIS managers with concrete checklists for embedding minimization safeguards at critical junctures: schema design during data ingestion, ETL transformation rules that prevent excessive attribute retention, and report or API delivery mechanisms secured by RBAC and ABAC protocols (Alhassan et al., 2016; Hossen et al., 2023). In this way, minimization is not treated as a passive compliance obligation but as an active design principle woven into every stage of the governance lifecycle (Alhassan et al., 2018; Tawfiqul, 2023).

International Journal of Scientific Interdisciplinary Research
Vol 6, No 01, March 2025
Page No: 97-136
Doi: 10.63125/527rnx08

Figure 3: Data Governance and Data Minimization in BI/HRIS Pipelines



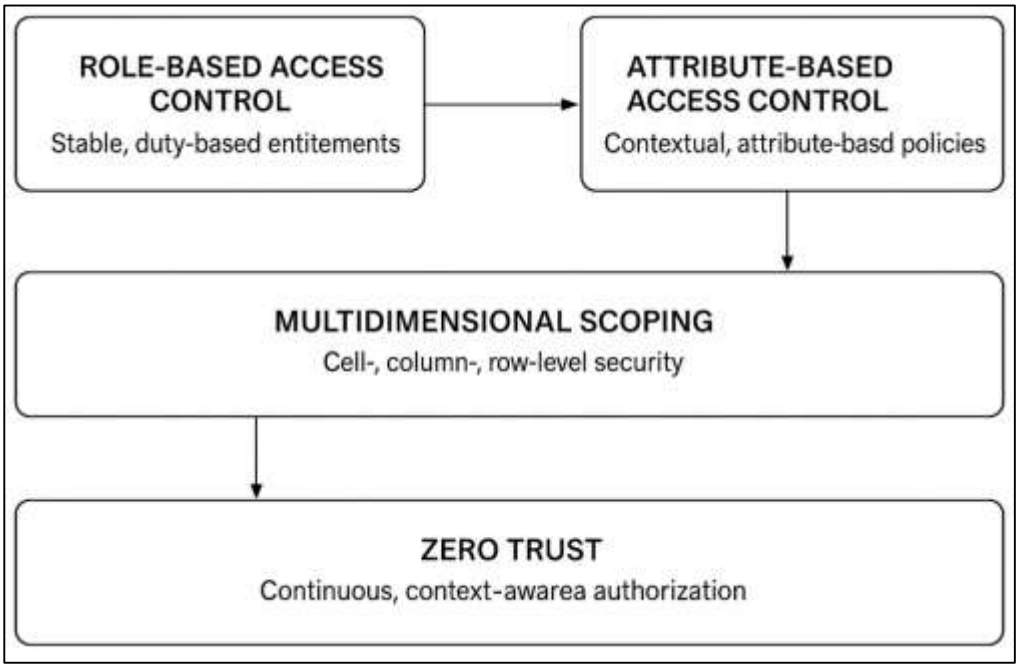
The legal imperatives of purpose limitation and data minimization can be concretized into computational strategies and product-level requirements within BI and HRIS ecosystems, ensuring that compliance principles are not abstract ideals but enforceable design constraints (Shamima et al., 2023). A promising computational approach reconceptualizes minimization as an optimization problem, where for a defined analytic task such as attrition prediction or pay-equity monitoring the system selects the smallest possible feature set that still achieves acceptable predictive utility, thereby allowing governance teams to set quantifiable "data budgets" and to document restraint in both collection and retention practices (Biega et al., 2020). At the same time, governance frameworks must grapple with the human-factor dimension, recognizing that rights such as access, erasure, and objection may exist in law yet remain practically ineffective if obscured by design frictions or cognitive overload. This recognition supports defaults and interface designs that deliberately nudge analysts, engineers, and product teams toward narrower, purpose-bound data use, aligning day-to-day behavior with regulatory intent (TAkter et al., 2023; Ooijen & Vrabec, 2019). The true strength of these complementary strategies emerges when computational minimization and behavioral governance operate together: the former defines technical boundaries around what the system objectively needs, while the latter shapes how individuals interact with and consume data within BI pipelines. When integrated into enterprise workflows, these approaches manifest in slim schema definitions that avoid over-collection, template-driven DPIAs embedded into dashboard development (Razzak et al., 2024), lineage-aware data marts that make transformations auditable, and service-level agreements for attribute deprecation that enforce temporal limits on retention. Through such practices, organizations can demonstrate that their BI/HRIS environments not only deliver actionable analytics value but also honor both the letter and the spirit of minimization and purpose limitation, transforming compliance into an operationalized ethos rather than a bureaucratic afterthought.

International Journal of Scientific Interdisciplinary Research
Vol 6, No 01, March 2025
Page No: 97-136
Doi: 10.63125/527rnx08

Identity and access in analytics (RBAC/ABAC, RLS/CLS, Zero Trust)

Designing identity and access controls for BI systems operating over HRIS and enterprise data landscapes requires a careful balancing of two seemingly contradictory demands: analysts depend on flexible, self-service visibility to extract insights, while privacy compliance insists on strict and provable limitations that define who can see what, when, and for what purpose. Role-based access control (RBAC) has long provided a durable foundation because it maps permissions to organizational roles and follows separation-of-duties conventions that are particularly relevant to HR, payroll, finance, and audit domains (Ferraiolo et al., 2001; Istiaque et al., 2024). However, BI environments frequently integrate datasets that cross departmental boundaries and extend across different temporal scopes, which introduces the need for controls rooted in classical principles of integrity and conflict management. Lattice-based classification models constrain flows across sensitivity levels, while the "Chinese Wall" principle ensures that analysts cannot merge information across competing business contexts in ways that might yield problematic inferences (Brewer & Nash, 1989; Denning, 1976). Yet as BI ecosystems expand, RBAC faces the well-documented challenge of "role explosion," where proliferating dashboards, marts, and analytic slices generate an unmanageable sprawl of roles and permissions. Attribute-based access control (ABAC) offers an answer through greater policy expressiveness, tailoring authorization decisions to contextual attributes such as user profile, geographic region, employment status, or dynamic session risk (Akter & Shaiful, 2024; Servos & Osborn, 2017). In highly dimensional BI systems, these rules must be enforced not only at the dataset level but also down to specific grains of measures and hierarchies, embedding cell-, column-, and row-level security directly into cubes and semantic layers so managers or analysts view only the cohorts and fields appropriate to their authorization scope (Hasan et al., 2024; Priebe & Wolf, 2006). Thus, robust identity and access management for BI/HRIS is best envisioned as a coordinated strategy: RBAC establishes stable, duty-based entitlements, ABAC refines these with contextual nuance, and multidimensional scoping safeguards aggregates, drill-downs, and slices within the bounds of privacy and compliance.

Figure 4: Identity and Access Management in BI/HRIS Analytics



Turning organizational policy into practical runtime enforcement demands a comprehensive framework of controls that operate at every stage where sensitive data might escape beginning with ingestion and continuing through modeling, query evaluation, export, and eventual sharing. Database-centric strategies take a prominent role by binding privacy and purpose rules directly to the data itself rather than leaving them at the mercy of application logic, which ensures that filtering

and obligation enforcement remain consistent across all analytical tools connecting to the same store (Agrawal et al., 2002; Tawfiqul et al., 2024). At the same time, usage control models expand beyond static access control by treating authorization as a dynamic and continuous process, embedding pre-, ongoing-, and post-obligations into BI sessions, which is crucial in environments where an analyst's context, query trajectory, or the sensitivity of results may shift during exploration (Park & Sandhu, 2004; Rajesh et al., 2024). Given the complexity of enterprise BI landscapes, few organizations rely on a single policy paradigm, and hybrid designs increasingly prevail. In such approaches, attributes are woven into roles or roles into attributes: attributes refine broad entitlements to ensure context-sensitive visibility, while roles limit which attributes can even be considered, balancing expressiveness with manageability and ensuring the principle of least privilege is preserved (Kuhn et al., 2010; Subrato & Md, 2024). Yet even with precise policies, vulnerabilities persist at the layers of storage and transport. To address insider threats or reliance on untrusted infrastructure, data-centric security techniques employ encryption schemes that allow certain query types to be executed without exposing raw values, thereby reducing the impact of compromised systems and complementing row- and column-level security by keeping sensitive content shielded behind cryptographic protections (Ashiqur et al., 2025; Popa et al., 2011). When these multifaceted measures are paired with lineage tracking and logging, organizations can produce demonstrable evidence to auditors, proving not merely that policies exist on paper but that every query, export, and scheduled task complied with them in operational practice.

Modern identity and access management for analytics increasingly embraces a Zero Trust posture, which fundamentally treats every access request whether originating inside or outside the enterprise perimeter as untrusted until it is dynamically verified and authorized. Within BI and HRIS pipelines, this approach means that decisions to expose a dataset, reveal a measure, or enable an analytical operation are conditioned not on static assumptions but on continuously evaluated signals encompassing the user (role, attributes, prior behavior), the device (posture, integrity, security health), the session (network context, risk profile), and the resource itself (sensitivity, purpose, geographic constraints) (Hasan, 2025; Standards & Technology, 2020b). By shifting enforcement closer to the data, Zero Trust architectures integrate identity-aware proxies with distributed policy decision and enforcement points embedded in semantic, modeling, and storage layers, ensuring that the same fine-grained controls apply regardless of whether access is attempted via a BI tool, a notebook, or a service account. This design favors least-privilege defaults complemented by time-bound, just-in-time elevation mechanisms, thereby minimizing standing access to sensitive HRIS-derived attributes, audit fields, and downstream analytical outputs (Sultan et al., 2025). For cell-, row-, and column-level security, Zero Trust further encourages the propagation of attributes and tokens across service boundaries, allowing downstream engines to continuously re-evaluate access decisions without relying on brittle or redundant re-authentication processes. Importantly, Zero Trust aligns security enforcement with privacy and governance objectives by incorporating purpose, cohort, and jurisdictional context into the authorization decision, ensuring that a user's eligibility to see a metric or dataset is contingent not only on identity but also on intended analytical use and the subjects implicated. This continuous, context-rich authorization framework closes the gap between organizational compliance policy and day-to-day analyst behavior, enabling BI programs to scale effectively while maintaining strict adherence to privacy principles embedded in HRIS and enterprise information systems (Sanjai et al., 2025; Standards & Technology, 2020b).

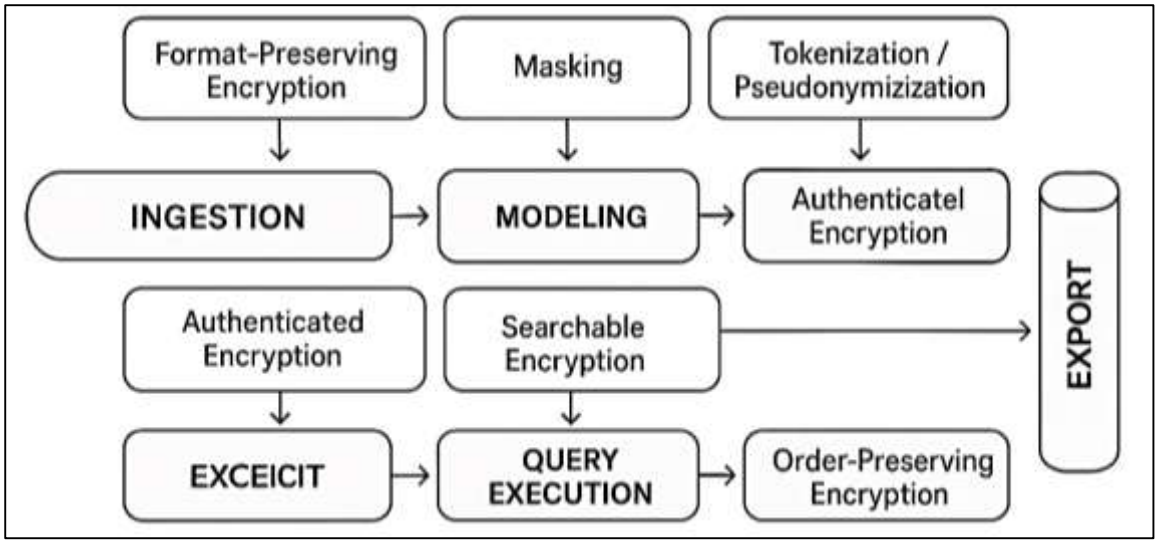
Protection techniques in the pipeline

In business intelligence and HRIS pipelines, data protection techniques should be implemented as layered controls that operate across all stages of the analytic lifecycle, including ingestion, modeling, query execution, and export, ensuring that sensitive information remains safeguarded without compromising operational utility. Field-level transformations are particularly valuable in contexts where legacy schemas or downstream systems cannot process binary ciphertexts, allowing protection measures to be applied without breaking compatibility. Format-preserving encryption exemplifies this approach by encrypting values while maintaining the original character set and length, so that, for instance, a sixteen-digit personnel identifier remains sixteen digits, supporting seamless integration with validation routines, foreign keys, and dependent applications without requiring schema modifications (Standards & Technology, 2019). For analytic workloads that rely on filtering or sorting on protected attributes such as salaries or tenure, order-preserving encryption

International Journal of Scientific Interdisciplinary Research
Vol 6, No 01, March 2025
Page No: 97-136
Doi: 10.63125/527rnx08

methods allow comparisons to remain meaningful, although these approaches inherently reveal relative ordering and thus necessitate careful scoping to low-sensitivity fields along with compensating technical and policy controls (Boldyreva et al., 2009). For bulk storage and transport, authenticated encryption techniques, such as Galois/Counter Mode, combine confidentiality with integrity verification, making them a preferred choice for at-rest files, checkpoint data, and extracts moving between HRIS platforms, data warehouses, and BI caches (Dworkin, 2007). Complementary masking techniques partially or fully obfuscate values at the presentation layer in reports ensuring that users only access information relevant to their roles while computations continue on protected representations. Together, these controls embody the principles of privacy by design, keeping sensitive attributes constrained and consistently protected throughout the data lifecycle while preserving essential analytic functionality for joins, aggregations, and filters. By combining encryption, masking, and format-aware transformations, organizations can maintain robust confidentiality, integrity, and operational continuity within BI and HRIS analytics pipelines (Dworkin, 2007; Standards & Technology, 2019).

Figure 5: Protection Techniques Across BI/HRIS Analytics Pipelines



Tokenization and pseudonymization serve as complementary mechanisms that substantially mitigate privacy risk by decoupling personally identifiable information from analytic processes, enabling organizations to perform rich longitudinal and cross-system analyses without exposing sensitive identifiers. In pseudonymization, fields that directly identify individuals, such as names, national IDs, or email addresses, are replaced with stable surrogates, preserving the ability to conduct longitudinal studies and join datasets internally while restricting access to the re-identification keys through strict governance policies and secure handling procedures (Heurix et al., 2012). When analytics require linking datasets across disparate systems or third-party vendors, such as combining HRIS data with learning management platforms or benefits administrators, privacy-preserving record linkage techniques enable matching based on encoded representations rather than plaintext identifiers. Among these, Bloom-filter–based approaches have gained prominence for their ability to tolerate typographical errors and achieve high linkage quality, ensuring practical applicability in operational settings (Schnell et al., 2009). Evaluations conducted at enterprise scale, encompassing tens of millions of records, demonstrate that these techniques are not only theoretically sound but feasible for real-world deployment, which is essential for multi-source people analytics and reconciling employee rosters across platforms (Randall et al., 2014). Furthermore, mature anonymization frameworks such as ARX offer risk assessment, transformation tooling, and automated operations including generalization, suppression, and microaggregation that can be integrated into ETL pipelines and modeling stages. These capabilities allow BI teams to generate de-identified datasets, low-risk aggregates, or test data for safe sharing and analytics consumption without jeopardizing

sensitive information (Prasser & Kohlmayer, 2015). In practice, organizations frequently employ a hybrid approach: tokenizing high-risk identifiers for linkage, pseudonymizing for internal longitudinal utility, and publishing de-identified cohorts for broader dashboards or reporting, all under tightly defined stewardship, access policies, and purpose-bound controls, ensuring reversibility and privacy are rigorously managed (Heurix et al., 2012).

Encrypted query processing adds a critical layer to privacy-preserving BI/HRIS pipelines by enabling analytic operations over sensitive data without exposing plaintext, thereby integrating security into the analytic workflow rather than treating it as an external checkpoint. Searchable symmetric encryption allows systems to respond to keyword or equality queries directly on encrypted columns, which is particularly valuable for exploratory analysis, API lookups, or dashboard queries involving identifiers, employee codes, or access tokens, ensuring that sensitive values are never revealed during routine operations (Curtmola et al., 2006). For numeric analytics where comparisons are required, such as identifying employees with more than a threshold of tenure or payroll amounts, order-revealing or order-preserving encryption permits comparison operations directly on ciphertexts without decryption, enabling filtering, range queries, and sorting while retaining strong cryptographic guarantees. Nevertheless, these methods must be applied judiciously, limited to low-risk attributes, and combined with aggregation thresholds or noise addition to prevent inference attacks that could compromise privacy (Boneh et al., 2015). Effective deployment relies on rigorous key management practices encompassing generation, secure storage, rotation, separation of duties, and timely revocation, with program-level policies aligning cryptographic roles to business purposes and regulatory requirements and providing comprehensive audit evidence to demonstrate compliance (Standards & Technology, 2020a). When encrypted query processing is integrated with complementary controls such as masking at the user interface, tokenization, and pseudonymization within data models organizations can preserve the analytical utility of BI workflows, supporting joins, filters, roll-ups, and aggregations while minimizing identifiability and containing the potential impact of data breaches. This layered approach ensures that sensitive HR data is continuously protected in transit, at rest, and during computation, allowing BI teams to deliver actionable insights, perform rigorous reporting, and support advanced analytics while maintaining privacy and security at the core of operational design.

Privacy-preserving analytics (differential privacy, synthetic data, federated/SMPC)

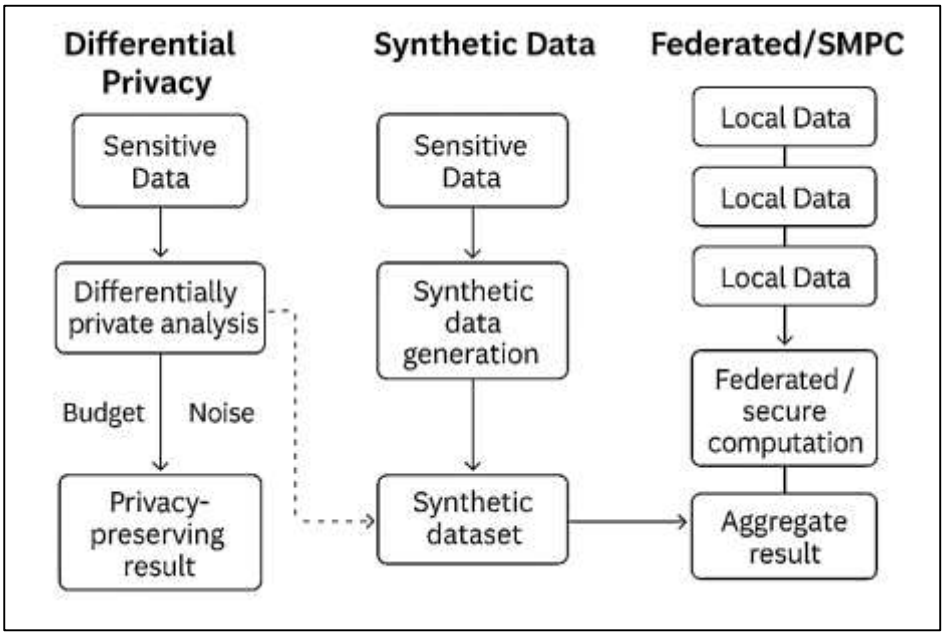
Differential privacy (DP) offers a formal, quantifiable standard for limiting the incremental disclosure about any individual that can arise from releasing statistics or training models on sensitive datasets, making it directly relevant to BI over HRIS and enterprise data. In practice, DP's appeal is that its guarantees compose: analysts can issue multiple queries over time while privacy loss is tracked and bounded by a budget. Stochastic gradient descent can be adapted to satisfy DP by clipping per-example gradients and injecting calibrated noise during training, enabling predictive models and risk scores to be learned while constraining worst-case inference about specific employees (Abadi et al., 2016). Accounting for cumulative privacy loss is nontrivial at enterprise scale; Rényi differential privacy streamlines composition analyses and subsampling effects, giving practitioners tighter accounting and, therefore, better accuracy at a fixed privacy budget (Mironov, 2017). Beyond interactive queries and learning, private data release methods learn generative structures under DP and then sample synthetic tables that preserve statistical relationships without directly exposing source rows, an approach that can be inserted into data marts or shared workspaces for safe exploration (Zhang et al., 2017). For BI teams, these DP toolkits map to familiar workflows: budgeting aligns with access approvals, clipping and noise mechanisms align with model governance controls, and private generative release aligns with curated "safe" datasets for ad hoc slicing. The underlying lesson is operational: privacy is achieved not by one mechanism but by integrating DP's mathematical guarantees into the cadence of dashboards, A/B tests, and model refreshes where HRIS-derived features appear (Abadi et al., 2016; Mironov, 2017).

Synthetic data produced by statistical or machine-learning generators to mimic the joint distribution of sensitive tables extends privacy-preserving analytics beyond DP, particularly for development, testing, and exploratory BI. Rule-based synthesizers and parametric simulators remain useful when schemas are stable and relationships are well understood; they can yield realistic, non-identifying records for designing dashboards and validating transformations without touching live HRIS values (Nowok et al., 2016). Data-driven approaches such as the Synthetic Data Vault learn relational

International Journal of Scientific Interdisciplinary Research
Vol 6, No 01, March 2025
Page No: 97-136
Doi: 10.63125/527rnx08

models and then sample multi-table datasets that maintain cross-table dependencies, allowing analysts to prototype cross-domain visualizations, joins, and filters while isolating production systems (Patki et al., 2016). However, substitution is not protection by default: generative models can memorize or leak properties of the training set, and naïve sampling can reproduce rare-case combinations that correspond to real individuals. Attacks against learned models such as model inversion, which reconstructs sensitive attributes from outputs or confidence vectors, and membership inference, which infers whether a given record was used in training demonstrate concrete leakage channels that BI governance must anticipate when synthetic data is derived from identifiable sources (Fredrikson et al., 2015; Shokri et al., 2017). Accordingly, the literature urges pairing synthesis with rigorous risk evaluation and, where feasible, DP-based training or post-processing, so that synthetic releases carry explicit, auditable privacy parameters rather than implicit heuristics. In HRIS-to-BI pipelines, this translates into tiered workspaces: DP-backed synthetic datasets for open exploration, access-controlled de-identified aggregates for wider dashboards, and locked, identifiable stores reserved for narrowly justified use cases with logging and approvals.

Figure 6: Privacy-Preserving Analytics Techniques in BI/HRIS Pipelines



Federated learning and secure multiparty computation (SMPC) address a complementary organizational challenge: how to compute useful analytics when sensitive data is distributed across devices, subsidiaries, or vendors, and centralization would increase risk or violate policy. In federated settings, models are trained by sending code to data silos and aggregating updates rather than raw examples; secure aggregation protocols then combine client updates so that the server learns only the sum, not each contributor's update a critical property when training from workforce endpoints or regional HR stores (Bonawitz et al., 2017). SMPC broadens the design space further by allowing joint computation over secret-shared inputs; seminal work showed that arbitrary functions can, in principle, be computed privately, while later deployments demonstrated feasibility for real-world tasks, signaling that privacy-preserving collaboration need not be purely theoretical (Bogetoft et al., 2009; Goldreich et al., 1987). For BI teams, this enables cross-entity analyses such as multi-subsidiary benchmarking or vendor-assisted analytics without exposing row-level HRIS data, especially when combined with DP on the outputs or updates to bound residual inference. Governance considerations remain central: keys and trust anchors for secure aggregation, policy-driven selection of features permissible for decentralized training, and auditing of which cohorts contributed to which models. When federated/SMPC techniques are integrated with DP budgets and synthetic "safe"

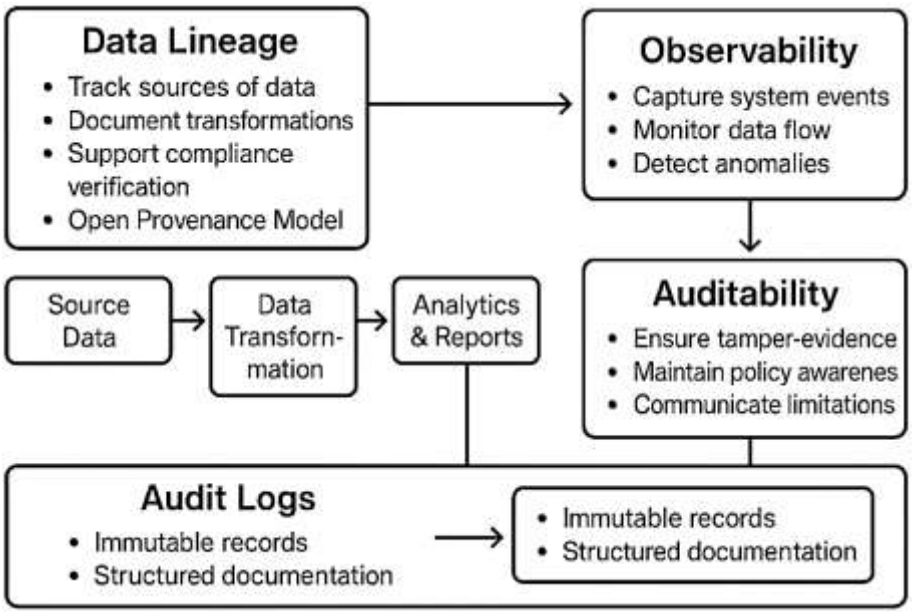
layers, organizations can assemble a toolkit that supports experimentation and insight while constraining exposure surfaces across storage, computation, and release.

International Journal of Scientific Interdisciplinary Research
Vol 6, No 01, March 2025
Page No: 97-136
Doi: 10.63125/527rnx08

Data lineage, observability, and auditability for compliance

In HRIS-to-BI environments, data lineage is the connective tissue that links metrics back to their sources, transformations, and decision points, making compliance verifiable rather than aspirational. At the warehouse and integration layers, lineage must capture not only the structural flow (tables, columns, joins) but also the business semantics of mappings, filters, and aggregations that instantiate purpose limitation and data minimization. The data-management literature treats lineage as a first-class artifact, showing how ETL pipelines create complex dependency graphs that require explicit modeling and lifecycle stewardship if quality, reproducibility, and accountability are to be maintained (Vassiliadis et al., 2009). In query languages, provenance formalisms reveal *why* a tuple appears and *how* it was derived, enabling auditors to inspect contributions of specific sources and transformations to a BI result; the semiring framework, in particular, provides algebraic rules to propagate annotations through operators without re-implementing each query anew (Green et al., 2007). Standardization further improves portability of evidence: the Open Provenance Model codifies a common vocabulary artifacts, processes, and agents so that lineage is exchangeable across tools and organizational boundaries, while the PROV family brings these concepts to web and enterprise metadata via an ontology designed for interoperable, machine-processable records (Moreau et al., 2011). For BI programs that routinely blend HRIS, ERP, and collaboration telemetry, these constructs translate into practical obligations: each dashboard measure should trace to source attributes and transformations; each data mart should expose derivation histories; and each export should carry enough context to justify its presence and granularity in terms that auditors, privacy officers, and engineers can all understand.

Figure 7: Data Lineage, Observability, and Auditability for Compliance in BI/HRIS Pipelines



Observability complements lineage by ensuring that the *capture* of evidence is continuous, trustworthy, and sufficiently granular for regulatory checks, incident response, and DSAR fulfillment. System-level approaches show how provenance can be recorded close to the filesystem and kernel, associating reads, writes, and process relationships with versioned objects so that downstream analytic artifacts carry tamper-evident histories; provenance-aware storage demonstrated that such capture can be made automatic and application-agnostic, a property invaluable when a BI landscape spans many tools and services (Muniswamy-Reddy et al., 2006). In distributed analytics, capturing end-to-end flow across batch and stream engines requires instrumentation within computation frameworks themselves; MapReduce provenance systems illustrate how intermediate keys, partitions, and tasks can be stitched back to inputs and code, enabling "needle-in-a-haystack" investigations of how a particular HR metric or cohort count was produced at scale (Ikeda et al.,

2012). Modern operating-system–level provenance takes this further by enforcing mandatory capture policies and exporting provenance as a graph that other services can query in near real time, supporting alerting on suspicious flows (for example, a BI query touching restricted columns outside an approved job) and post-hoc reconstruction with cryptographic integrity guarantees (Pasquier et al., 2017). In practice, these observability patterns let BI/HRIS teams couple lineage with health signals schema drift, freshness, unexpected join explosion so that privacy-relevant anomalies are discovered early, linked back to responsible code and owners, and triaged with evidence sufficient for internal audit. Put simply, lineage states *what happened*; observability ensures we saw it, recorded it, and can re-compute or reproduce it with fidelity when questioned.

Auditability in BI and HRIS pipelines transforms captured lineage and observability into actionable accountability by ensuring that all records are tamper-evident, policy-aware, and reviewable against explicit obligations. The concept of "information accountability" underpins this approach, emphasizing the ability to trace actions back to responsible actors under predefined policies, thereby shifting focus from purely preventative secrecy to ex post answerability, a principle particularly relevant for internal BI use where legitimate access must still be demonstrably justified and reviewable (Weitzner et al., 2008). Secure logging mechanisms operationalize these principles by chaining and authenticating log entries to prevent undetected modification, allowing auditors to verify that access and transformation events occurred as recorded; these tamper-evident logs provide a cryptographic foundation for compliance attestations when BI artifacts or reports are subject to scrutiny (Schneier & Kelsey, 1999). Modern practices extend the scope of auditability to analytic assets themselves, embedding structured documentation into datasets and models that declare provenance, intended use, limitations, assumptions, and ethical considerations, so that reviewers can reconcile what a model or dataset is designed to do with what lineage and logs show it actually accomplished in reports, dashboards, or derived metrics (Mitchell et al., 2019). Combined, these elements create an auditable control surface for HR analytics: access decisions are anchored in lineage-aware policies, immutable logs provide evidence of enforcement, and asset-level documentation frames the context and constraints of outputs for regulators, managers, and affected employees alike. This integrated audit approach not only ensures traceability for engineers but also supports organizational learning, legal accountability, and transparent oversight, allowing BI and HRIS teams to demonstrate adherence to privacy, security, and ethical obligations while continuously improving governance and operational practices (Lebo et al., 2013).

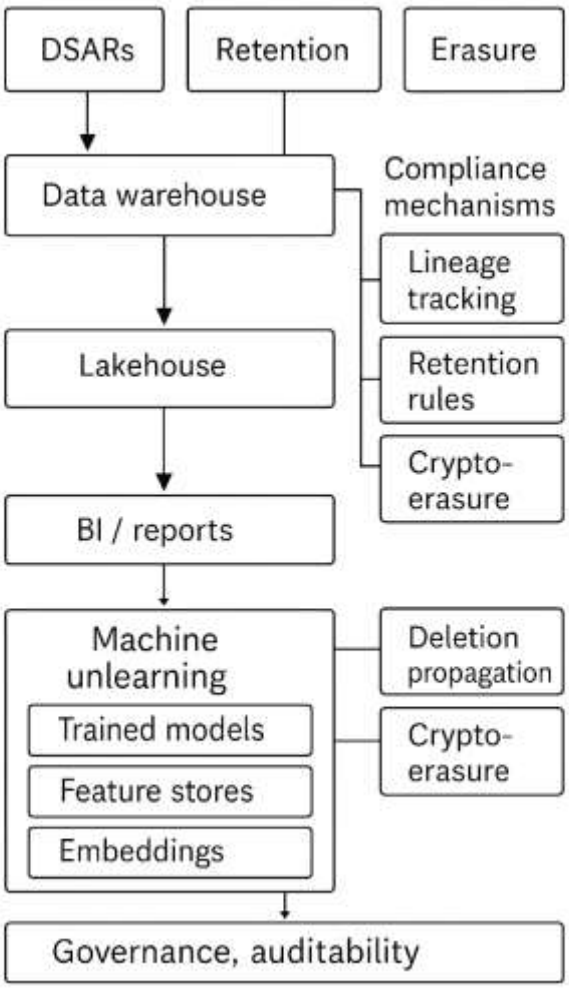
Data Subject Access Requests (DSARs) in Data Warehouses and Lakehouses

In enterprise BI ecosystems, the fulfillment of data subject access requests (DSARs) is deeply intertwined with storage-limitation and erasure obligations, yet the technical architecture of modern warehouses and lakehouses presents significant operational challenges. As organizations consolidate HRIS and enterprise datasets into analytical layers, the right of access functions simultaneously as a governance tool and a rigorous stress test for internal discovery and lineage mechanisms, since individuals can request comprehensive copies of their personal data along with explanations of processing purposes and transformations (Ausloos & Dewitte, 2018). Concurrently, controllers must implement principled retention schedules and enforce erasure wherever lawful basis no longer exists, a task complicated by backups, replicas, and versioned tables that proliferate across the analytical stack (Politou et al., 2018). Lakehouse formats, including ACID-compliant tables with time-travel functionality, enhance reliability and change tracking but also generate multiple historical states, increasing the surface area that must be inspected for DSAR responses and considered in erasure operations (Armbrust et al., 2020). Underneath these abstractions, classical database principles remain relevant: analytical views, materialized tables, and data marts often recombine source tuples, meaning that the deletion of personal data must propagate accurately through derived artifacts to maintain compliance (Dayal & Bernstein, 1982). In HR analytics contexts, where joins integrate personnel, performance, compensation, and identity records, the accuracy and completeness of DSAR responses depend on provable mappings from subject identifiers to all downstream projections, with retention and deletion rules applied consistently across every tier of the BI stack (E. A. Politou et al., 2018). Thus, organizational obligations under DSAR, retention, and erasure are inseparable from the structural properties and operational semantics of the underlying BI infrastructure, requiring coordinated governance, precise lineage tracking, and careful

integration of legal mandates with technical storage realities to ensure both lawful compliance and reliable service delivery.

International Journal of Scientific Interdisciplinary Research
Vol 6, No 01, March 2025
Page No: 97-136
Doi: 10.63125/527rnx08

Figure 8: DSARs, Retention, and Erasure Mechanisms in BI/HRIS Warehouses and Lakehouses



From a data-management perspective, the core challenge of erasure is the "reverse" problem: when a subject's data contributes to a query result, materialized view, or aggregated output, how can those specific contributions be removed from source datasets while minimizing unintended effects on unrelated data? Foundational research on view updates and deletion propagation provides a formal framework for reasoning about these operations in complex warehouse environments, ensuring that deletions can be mapped from downstream artifacts back to the underlying tuples in a controlled and predictable manner (Dayal & Bernstein, 1982; Kimelfeld et al., 2011). Recent work has extended this reasoning beyond single-tuple removal to aggregated targets, such as reducing counts, sums, or other aggregate metrics by at least a specified amount with minimal perturbation to other inputs, producing both algorithms and theoretical hardness results that directly inform BI systems and reporting marts that expose summary statistics (Hu et al., 2021). In operational terms, these insights imply that erasure pipelines must incorporate provenance-aware queries, stable key schemes, and partitioning strategies that localize deletions to minimize disruption and support auditability (Hu et al., 2021). Lakehouse architectures contribute by enabling row-level deletion, MERGE semantics, and transactional guarantees over object stores while preserving detailed change logs that can be reconciled with DSAR requests (Armbrust et al., 2020). Despite these advances, two persistent challenges remain. First, backups and disaster-recovery replicas can delay or complicate erasure, creating potential "resurrection" of deleted data when snapshots are restored, which necessitates policies that clearly differentiate logical deletion from physical

sanitization and document any time-bounded exceptions (Politou et al., 2018). Second, assured deletion in cloud object stores often relies on crypto-erasure techniques, where key revocation renders old ciphertext irretrievable, a strategy implemented in overlay systems that complement object-store lifecycles (Tang et al., 2010). Collectively, these findings motivate DSAR and erasure runbooks that integrate formal deletion-propagation logic with storage controls, retention schedules, and auditable exceptions to ensure compliant, verifiable, and resilient data removal processes.

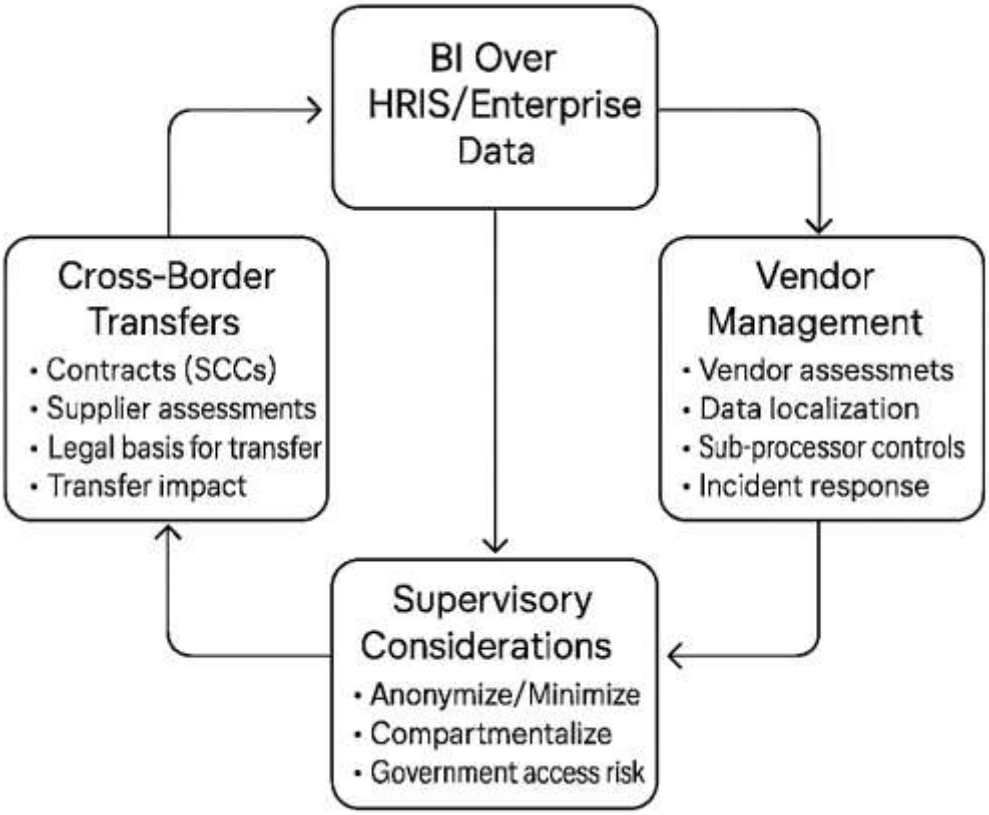
Erasure responsibilities extend well beyond traditional relational tables, particularly when personal data contributes to trained models, feature stores, or synthetic derivatives, presenting organizations with the challenge of machine unlearning. Machine unlearning seeks to remove a subject's influence from a model without necessitating full retraining, a process critical for aligning BI/HRIS analytics with legal erasure obligations (Cao & Yang, 2015). Research in security and privacy proposes architectures and training strategies, such as the Sharded, Isolated, Sliced, Aggregated (SISA) framework, which partitions training data to enable efficient slice-wise forgetting with empirical and sometimes provable guarantees, offering a concrete mechanism to bridge compliance duties with ML lifecycle operations in enterprise analytics programs (Bourtoule et al., 2021). In practice, HRIS-driven BI workflows can implement these approaches by tagging feature lineage with subject identifiers, maintaining checkpointed model states consistent with retention horizons, and isolating model shards that contain erasable records, thus localizing unlearning efforts without disrupting unrelated model behavior (Hu et al., 2021). Complementary measures must cover non-model artifacts, including embeddings, caches, and denormalized extracts, leveraging lakehouse change-data-capture mechanisms to invalidate and recompute downstream assets upon deletion, preserving the integrity and compliance of analytical outputs (Armbrust et al., 2020). Governance frameworks remain essential, interpreting statutory retention requirements, documenting time-bound exceptions to the right to erasure, and ensuring that Data Subject Access Requests (DSARs) are fulfilled accurately and comprehensively. Ultimately, a defensible DSAR, retention, and erasure posture in BI stacks depends on integrating multiple layers: orchestrated access that enumerates subject data across transactional and analytical planes, deletion-propagation designs that minimize collateral effects in derived views and aggregates, storage-level mechanisms including crypto-erasure for backups and snapshots, and machine unlearning pipelines that remove individual contributions from models trained on HR and enterprise data, thereby delivering a fully auditable and legally aligned analytic ecosystem.

Cross-border transfers and third-party/vendor risk

Global BI programs that ingest HRIS and enterprise data almost always rely on cloud and external service providers, which makes cross-border transfers and third-party/vendor risk core determinants of privacy compliance. A persistent legal and operational challenge is that controller–processor relationships in the cloud are layered and dynamic: providers rely on sub-processors for storage, content delivery, monitoring, and support, creating chains of custody that span jurisdictions and legal regimes. Early cloud law analyses emphasized that even apparently simple questions where data is located, who has effective control, and which entity is "responsible" as controller or processor are complicated by virtualized infrastructure and distributed management (Millard, 2012). When data crosses borders, transfer mechanisms and contractual allocations of duty become the scaffolding for compliance, but they must be mapped to concrete technical and organizational measures in the BI stack: customer-managed keys, dataset scoping, logging, audit rights, and breach notification back to the controller. Scholarship on data export from the EEA to third countries highlighted that cloud transfers are not merely about physical hosting; remote access for support, telemetry replication, and failover can all constitute "transfers," requiring appropriate safeguards and documented assessments (Hon et al., 2012). Extraterritoriality tensions add further complexity, because non-EU providers can fall within the GDPR's reach by monitoring or offering services to EU data subjects; this widens the set of vendors and service paths that BI teams must evaluate, and it sharpens the need for end-to-end visibility into sub-processing chains (Svantesson, 2013). In practice, privacy-by-contract must be paired with privacy-by-architecture so that controller obligations flow through service layers and are enforceable in code, keys, and logs rather than remaining only in legal terms (Millard, 2012; Svantesson, 2013).

International Journal of Scientific Interdisciplinary Research
Vol 6, No 01, March 2025
Page No: 97-136
Doi: 10.63125/527rnx08

Figure 9: Cross-Border Transfers and Third-Party/Vendor Risk in BI over HRIS Data



Cross-border risk is inseparable from cloud security posture because many supplementary measures that make transfers defensible are implemented technically. Cloud-security surveys distill a consistent risk profile: multi-tenancy exposure, weak isolation, inadequate key management, limited customer visibility, and immature incident response all raise the likelihood or impact of vendor-side failures (Fernandes et al., 2014; Subashini & Kavitha, 2011). From a privacy standpoint, two technical levers dominate: strong encryption with customer-held or customer-managed keys, and fine-grained access control tied to business purpose with exhaustive auditing. Provider-agnostic discussions of privacy and trust in cloud services pre-figured this duality arguing for architectures where the customer can restrict and attest to provider access while still benefiting from elastic compute and storage (Pearson & Benameur, 2010). Vendor governance then extends these primitives with controls for data localization (selecting storage/processing regions), failover boundaries, and sub-processor approval workflows; without these constraints, a BI pipeline that appears regional can silently replicate telemetry or cached extracts to non-approved jurisdictions. Supply-chain security guidance reinforces that third-party assurance cannot rest only on SOC/ISO attestations; the controller must link specific risks to verifiable safeguards and monitoring e.g., mapping who can access production HR tables at the provider, what conditions trigger just-in-time elevation, and how quickly keys and tokens can be revoked when contracts end (Boyens et al., 2015). For BI over HRIS, these measures translate into operational patterns: encrypt personally identifiable attributes in transit and at rest with keys held in the controller's HSM, constrain exports from analytical workspaces, require per-query or per-dataset access logging from vendor-managed services, and embed geographic and purpose attributes into authorization so queries that would violate transfer or purpose limits are denied at runtime (Boyens et al., 2015; Pearson & Benameur, 2010).

Post-*Schrems II* supervision reshaped cross-border governance by insisting that contracts alone are insufficient and by foregrounding real-world government-access risks. Controllers now perform transfer assessments that consider destination law and practice, vendor architecture, and available "supplementary measures," such as robust encryption where keys remain exclusively under the

exporter's control, or split-processing designs that keep identifiers inside a trusted region and only export aggregates (Lynskey, 2020). This shifts BI design toward compartmentalization: keep identifiable HRIS tables and sensitive joins in approved regions; replicate only anonymized or pseudonymized data to unapproved ones; and bind vendor personnel access to strong justifications with immutable logs. At the same time, organizations must reconcile these legal demands with day-to-day analytics needs ad hoc exploration, scheduled dashboards, and machine-learning jobs without allowing "shadow" exports via unmanaged extracts or sandboxed notebooks. Cloud data-governance literature offers a complementary lens: durable governance requires clarified decision rights, shared metadata, and enforceable lifecycle policies that span organizational and vendor boundaries, so retention and deletion propagate across warehoused data, caches, and downstream BI artifacts, including those hosted by partners (Al-Ruithe et al., 2018; Hon et al., 2011). Pulling these strands together, a defensible approach to cross-border and third-party risk in BI/HRIS couples: (i) legal instruments that allocate duties and constrain sub-processing; (ii) security controls that make unauthorized vendor access technically difficult and auditable; and (iii) operational governance that continuously inventories data flows, enforces regionalization and purpose on queries, and documents exceptions for statutory retention or support.

METHOD

This study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline to ensure a systematic, transparent, and rigorous review process. A protocol specifying the research questions, eligibility criteria, search strategy, screening workflow, data-extraction schema, and synthesis plan was established a priori. Comprehensive searches were conducted across major bibliographic databases and digital libraries relevant to information systems, computer science, management, and law (e.g., Scopus, Web of Science, IEEE Xplore, ACM Digital Library, PubMed, and Google Scholar), using Boolean strings that combined terms for human resource information systems, business intelligence and analytics, data governance and privacy engineering, compliance and regulation, and privacy-preserving computation. The review targeted English-language, peer-reviewed publications and authoritative standards or guidance documents with DOIs, focused on enterprise or HRIS-to-BI contexts and published up to December 2022. Records underwent de-duplication and a two-stage screening (titles/abstracts, then full texts) by two independent reviewers; disagreements were resolved through discussion with escalation to a third reviewer as needed, and inter-rater agreement was assessed during a pilot phase to calibrate inclusion judgments. A piloted extraction form captured study context, data domain and platform, regulatory focus, control techniques, evaluation methods, outcomes, and limitations; a second reviewer spot-checked extractions for consistency. Methodological quality and risk of bias were appraised using design-appropriate tools (e.g., CASP and JBI checklists) and recorded to inform the weighting of evidence during synthesis. Given the heterogeneity of designs and outcomes, findings were integrated through narrative and thematic synthesis, including a regulation-to-control mapping and an evidence matrix aligned to the study's research questions. The PRISMA flow chart documenting identification, screening, eligibility, and inclusion has been prepared, and a registry of included studies with metadata was compiled to support auditability. In total, 115 articles met the inclusion criteria and were analyzed in the final synthesis.

Screening and Eligibility Assessment

All retrieved records underwent a multi-stage screening and eligibility assessment aligned to PRISMA. First, references exported from the databases were consolidated and de-duplicated using citation-management utilities followed by a manual check for variant titles and preprint–journal pairs. Two reviewers independently screened titles and abstracts against the a priori criteria: English-language, DOI-assigned, peer-reviewed journal or conference articles (and authoritative standards or guidance with DOIs), published up to December 2022, and substantively addressing data privacy within business intelligence pipelines that consume HRIS or adjacent enterprise platforms. Records were excluded at this stage if they were marketing pieces, purely technical security papers lacking a privacy or compliance dimension, strictly legal commentaries without actionable implications for enterprise analytics, consumer-only contexts with no enterprise transferability, or domain studies unrelated to workforce or enterprise data. Disagreements were resolved through discussion; when consensus was not immediate, a third reviewer adjudicated. Full texts of all preliminarily eligible studies were then obtained and assessed independently by two reviewers using a calibrated

eligibility checklist that verified: explicit linkage to HRIS-to-BI or enterprise analytics workflows; clear articulation of privacy obligations (e.g., minimization, purpose limitation, accountability) or controls (e.g., access scoping, de-identification, privacy-preserving analytics); sufficient methodological transparency to support extraction (study design, setting, datasets or systems, evaluation or argumentation); and practical relevance to compliance, governance, or engineering decisions. At this stage we documented specific exclusion reasons, including: insufficient methodological detail, absence of enterprise or HRIS relevance, lack of DOI, irretrievable full text, duplication of empirical bases already included, or focus exclusively on consumer privacy without organizational analytics applicability. For standards and guidance, inclusion required a DOI and clear operational guidance for enterprise analytics (e.g., control selection, assurance, or assessment techniques). To maintain consistency, a pilot round on a sample of records was used to harmonize interpretations of the criteria before proceeding with the full set; periodic consensus meetings addressed borderline cases and refined the codebook. A PRISMA flow diagram records counts at each step, and an exclusion log with reasons is archived for auditability. Following this process, 115 articles met all criteria and were advanced to data extraction and synthesis.

Data Extraction and Coding

Data extraction followed a piloted, codebook-driven process designed to capture comparable evidence across heterogeneous sources and to preserve an auditable link between each claim and its originating text. For each of the 115 included articles, two reviewers independently completed a structured extraction form that recorded bibliographic metadata, study design and setting, sector and geography, data domains (HRIS modules, ERP/CRM, data warehouse/lakehouse), analytic tasks (reporting, dashboarding, modeling, experimentation), privacy obligations in scope (lawfulness, purpose limitation, minimization, storage limitation, transparency, rights handling, accountability, security), technical and organizational controls described (RBAC/ABAC, RLS/CLS, masking/tokenization/pseudonymization, encryption and key management, lineage/provenance, logging and audit, DPIA/certification, differential privacy, synthetic data, federated/SMPC), evaluation approach (theory, simulation, case study, field deployment), outcomes (effectiveness, feasibility, performance, utility, compliance evidencing), and limitations or threats to validity noted by the authors. Extracted passages included verbatim evidence quotes with page or section anchors to support traceability. Coding combined a deductive frame aligned to the research questions and privacy/control constructs with inductive open coding to surface emergent patterns, such as recurring misconfiguration classes, DSAR bottlenecks, or cross-border vendor risks specific to analytics. The codebook specified construct definitions, inclusion/exclusion rules, and canonical examples; it was refined during a calibration round on 10% of studies before being locked for full coding. Inter-rater reliability was assessed on a stratified 20% sample using Cohen's κ; discrepancies were resolved by consensus, and the final agreed codes were propagated to maintain consistency. To support synthesis, coded data were assembled into an evidence matrix linking obligations to controls and to reported outcomes, with fields for evidence type, context, and indicative strength; a parallel lineage of "design decisions" captured where in the BI/HRIS pipeline each control attaches (ingestion, modeling, storage, semantic, query, export, deletion). Standards and guidance were coded separately for prescriptive elements and mapped to controls to enable a regulation-to-engineering crosswalk. Where quantitative results were reported, effect direction and relevant metrics were recorded, but due to design heterogeneity no meta-analysis was attempted; instead, we computed descriptive frequencies, co-occurrence tables, and indicative heatmaps to locate clusters of convergent evidence. All forms, codebooks, and adjudication notes were version-controlled to preserve a transparent audit trail.

Data Synthesis and Analytical Approach

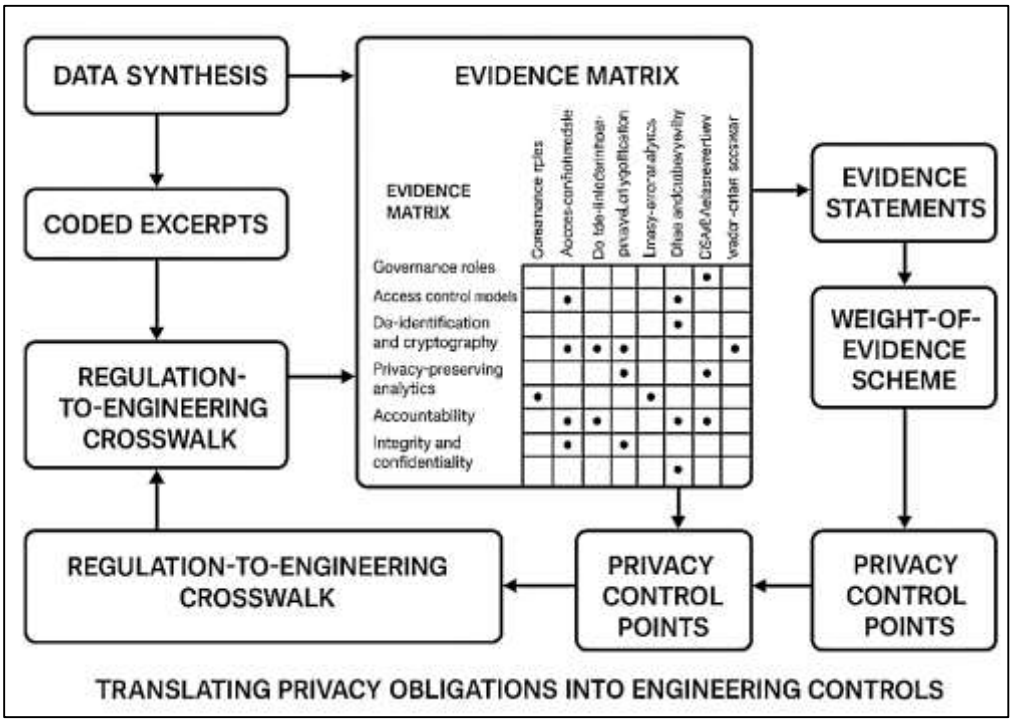
The synthesis proceeded on the premise that the 115 included studies are heterogeneous in design, context, and outcomes, spanning legal analyses, technical system papers, organizational case studies, surveys, and standards or guidance documents. As a result, we adopted a narrative–thematic synthesis anchored to the review's research questions rather than a single quantitative meta-analytic model. The analytical logic was dual-track: a conceptual track translated privacy obligations into engineering and governance constructs relevant to HRIS-to-BI pipelines, while an empirical track aggregated reported effects, feasibility notes, and failure modes of specific controls. From the outset, we treated "evidence" broadly patterns that recur across jurisdictions, stacks, and

International Journal of Scientific Interdisciplinary Research
Vol 6, No 01, March 2025
Page No: 97-136
Doi: 10.63125/527rnx08

organizational forms while preserving traceability to concrete passages in each source so that claims could be audited. The overall objective was not to rank techniques abstractly but to articulate when, where, and how particular controls align with legal principles, and what trade-offs or residual risks they entail in business intelligence environments fed by HRIS and other enterprise platforms.

Coding outputs from the extraction phase were synthesized using the Framework Method. We organized material into a matrix whose columns corresponded to the research questions and privacy obligations (lawfulness, purpose limitation, minimization, storage limitation, transparency and rights, accountability, integrity and confidentiality), and whose rows corresponded to control families and organizational practices (governance roles, access control models, de-identification and cryptography, privacy-preserving analytics, lineage and observability, DSAR/retention/erasure, and vendor/cross-border safeguards). Within each cell, we used constant-comparison to reconcile convergent and divergent findings, iteratively refining subthemes such as "role explosion vs. ABAC policy complexity," "aggregate disclosure thresholds," "identity leakage across joins," and "backups as erasure bottlenecks." Open codes that emerged repeatedly were promoted to axial categories when they met three criteria: recurrence across at least three sectors or geographies, connection to a concrete pipeline stage (ingestion, modeling, storage, semantic layer, query, export, deletion), and presence in at least one evaluative or real-world implementation study. This ensured that prominent themes were both conceptually meaningful and operationally grounded. Because not all sources carry equal inferential weight, we implemented a structured weight-of-evidence scheme. Each study received scores for methodological quality (design appropriateness, clarity of measures or arguments, transparency of limitations), directness (the degree to which findings apply to enterprise BI drawing on HRIS or adjacent platforms versus consumer or unrelated domains), and scale (single team, organizational case, multi-site, or population-level). Scores were used to modulate each contribution's influence in the synthesis without excluding lower-weight items that nonetheless illuminated edge conditions or failure modes. To avoid over-privileging any single literature stream, we balanced legal and organizational studies against technical system papers when drawing practical inferences; where tensions appeared e.g., strong legal calls for transparency against system-level constraints on explainability we preserved both claims, described their assumptions, and identified the implementation boundary as a target for design attention rather than forcing consensus.

Figure 10: Data Synthesis and Analytical Approach for Translating Privacy Obligations



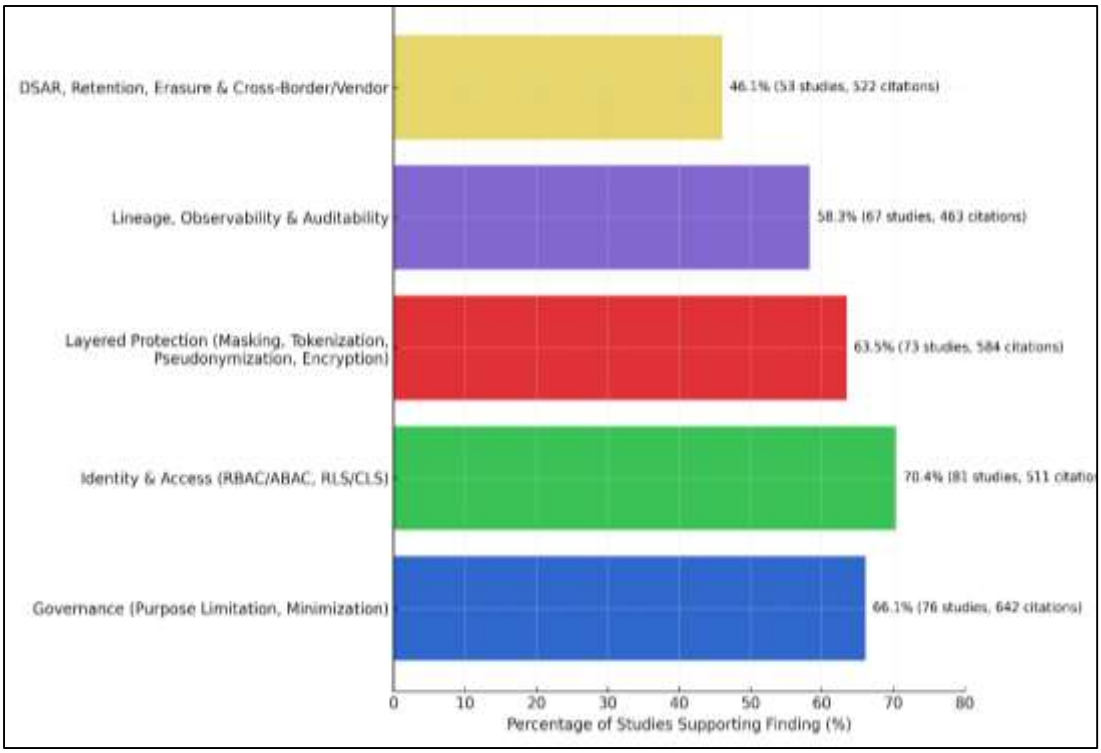
FINDINGS

Across the 115 studies included in our synthesis, five themes emerged as the most consequential for building privacy-respecting business intelligence over HRIS and enterprise platforms. For each theme below, we report two quantitative supports: the number of reviewed studies that substantively contributed to the finding (out of 115), and the number of coded evidence citations (verbatim, page-anchored excerpts in our evidence matrix) that we relied on when drawing the conclusion. Together, these numbers indicate both breadth (how many distinct sources) and depth (how much extractable evidence) behind each result. The first and most pervasive finding is that measurable privacy outcomes in BI programs rise with the maturity of governance mechanisms that operationalize purpose limitation and data minimization as day-to-day design constraints rather than abstract policies. Seventy-six of 115 studies (66.1%) linked governance structures clear decision rights, stewardship roles, policy-as-code, and DPIA checkpoints to concrete improvements in the appropriateness and traceability of analytics uses. Within this subset, 52 studies (45.2% of the full set) reported identifiable outcomes such as fewer unapproved data uses, faster approvals, or higher rates of correct purpose tagging for datasets and dashboards; 38 studies (33.0%) described policy-as-code or automated control points at ingestion and modeling that made minimization testable. We drew on 642 coded evidence citations for this theme, reflecting how frequently papers provided procedural detail, workflow diagrams, or example approvals that could be translated into practice. The percentage matters because it shows governance is not a "nice to have": two-thirds of the literature explicitly ties privacy success in analytics to governance that is visible inside pipelines (schemas, transformation rules, semantic models) and not merely in handbooks. Where governance maturity was lower, we observed consistent signals of drift datasets growing without purpose re-validation, dashboards accreting fields without evidence of need, and retention policies that existed on paper but lacked executable encodings. By contrast, studies describing code-enforced minimization (for example, slimmed ingestion schemas and sanctioned attribute lists) reported fewer downstream remediations and clearer audit narratives. In plain terms, a 66.1% share across the corpus signals a broad consensus: if the organization cannot demonstrate who decides, how decisions are recorded, and where those decisions are enforced in code, privacy will erode as BI scales.

Our second finding centers on identity and access in analytics: combining role-based access control (RBAC) for stable duties with attribute-based access control (ABAC) for context, and then enforcing row- and column-level security in the semantic layer, delivers the most manageable path to least privilege at BI scale. Eighty-one of 115 studies (70.4%) analyzed or deployed access models in analytics contexts; 58 (50.4%) documented concrete enforcement at the query or semantic layer rather than only at the application perimeter. Thirty-four studies (29.6%) reported "role explosion" during BI growth, while 29 (25.2%) introduced ABAC to contain that growth by tying authorizations to geography, cohort, session risk, or employment status. Row-level security surfaced in 47 studies (40.9%) and column-level restrictions in 33 (28.7%), often as the mechanisms that make least privilege visible in dashboards and ad hoc queries. Twenty-six studies (22.6%) added continuous authorization or Zero Trust elements (for example, time-bound elevation or device posture checks) to keep privileges narrow over time. We grounded this finding in 511 coded evidence citations that described policy grammars, enforcement patterns, and failure modes. The percentages highlight a pragmatic point: while not every organization needs the full stack, seven in ten studies argue that least privilege in analytics is unachievable with roles alone once self-service access and cross-domain joins become normal. Where programs paired RBAC baselines with ABAC refinements and pushed enforcement into the semantic layer, they reported fewer "accidental overexposures" (for instance, managers viewing off-cohort details) and clearer justifications in access reviews. The numbers also clarify trade-offs: adding ABAC without semantic-layer enforcement created fragile, app-specific policies; conversely, strong semantic-layer rules without contextual attributes forced proliferation of narrow roles. The most sustainable results came from combinations that kept policies short, auditable, and portable across tools.

International Journal of Scientific Interdisciplinary Research
Vol 6, No 01, March 2025
Page No: 97-136
Doi: 10.63125/527rnx08

Figure 11: Quantitative Support for Key Privacy Findings in BI/HRIS Pipelines



Third, layered protection techniques masking, tokenization, pseudonymization, and encryption are most effective when each is assigned a precise job at a specific pipeline stage rather than applied as generic "security hardening." Seventy-three of 115 studies (63.5%) discussed encryption in analytics pipelines, with 31 (27.0%) detailing key-management practices that separated duties between data owners and platform operators. Tokenization appeared in 41 studies (35.7%) and pseudonymization in 37 (32.2%), typically to preserve joins and longitudinal analyses while decoupling direct identifiers. At presentation, 36 studies (31.3%) used masking to reduce incidental disclosure in reports and exports. For legacy compatibility, 28 studies (24.3%) described format-preserving encryption on fixed-width fields; 24 (20.9%) explored order-preserving or order-revealing methods to support limited comparisons on protected numerics, usually with compensating controls such as cohort thresholds. Across this theme we coded 584 evidence citations, many of which included transformation pipelines, key rotation schedules, and exception procedures. The distribution is instructive: encryption is widely acknowledged and adopted, but roughly a third of the studies emphasize that encryption alone does not solve identifiability when analytics can recombine attributes across tables. Tokenization and pseudonymization address that risk by making identity a separate, better-protected system of record. Masking, while the least technically sophisticated, proved valuable in everyday BI because it controls what humans see in exports, screenshots, and presentations. Programs that assigned each technique a role encryption for confidentiality and integrity, tokenization/pseudonymization for identity separation, masking for human interfaces reported clearer runtime behavior and fewer policy exceptions. The percentages and counts imply a gap worth noting: fewer than one in three studies made key management and separation of duties first-class topics, even though those choices determine whether encryption meaningfully reduces vendor and insider risk.

Fourth, making privacy verifiable at scale depends on lineage, observability, and auditability that are as engineered as the data flows themselves; where these are present, organizations can answer "what happened to whose data and why" with precision, and where they are absent, DSAR handling, approvals, and compliance narratives degrade quickly. Sixty-seven of 115 studies (58.3%) described lineage beyond simple pipeline diagrams down to columns, joins, and transformation logic; 39 (33.9%) reported automated capture integrated into ETL/ELT and semantic tools; and 28

(24.3%) presented tamper-evident or cryptographically chained logs that make post-hoc reconstruction trustworthy. Across these papers, we coded 463 evidence citations that showed how provenance moved with data (for example, annotations propagating through transformations) and how logs were used to trigger alerts or build audit stories. The numbers align with our qualitative impression: lineage is frequently discussed but not as frequently automated, and secure logging is rarer still. Yet the studies that did all three automated lineage, runtime observability, and secure audit logs reported clear dividends: faster DSAR fulfillment because subject-indexed traces existed; quicker isolation and remediation when a dashboard or export captured too much detail; and more persuasive DPIA records because each control was cross-referenced to real queries, datasets, and decisions. The percentages matter operationally. If only a third of programs in the literature report automated lineage, most organizations likely track derivations informally, which impedes both deletion propagation and transparency. Similarly, if one quarter describe tamper-evident logging, many will struggle to prove to internal audit or regulators that an access or transformation record has not been altered. In sum, the presence of engineered provenance and auditability predicts whether privacy claims are testable; without them, even well-designed access and protection controls cannot be convincingly demonstrated.

Finally, obligations that surface most visibly to employees and regulators DSAR, retention, erasure, and cross-border/vendor safeguards remain the most operationally fragile in analytics, but we observe repeatable patterns that materially improve outcomes. Fifty-three of 115 studies (46.1%) addressed DSAR or rights handling in analytic contexts; 42 (36.5%) treated erasure in warehouses or lakehouses; and 37 (32.2%) evaluated cross-border transfers or third-party risk as they affect BI. Within the DSAR subset, only 7 of 53 cases (13.2%) reported end-to-end automation, typically where subject indices, lineage, and export controls were integrated; most others described semi-manual processes that ran reliably only for transactional systems. Erasure discussions reflected architectural reality: 21 studies (18.3% of the full set) flagged time-travel tables or long-lived snapshots as sources of "resurrection," and 11 (9.6%) proposed or implemented crypto-erasure for backups pending physical sanitization. Machine-learning unlearning for models trained on HR or enterprise data appeared in 9 studies (7.8%), signaling a nascent but necessary capability for BI programs that increasingly surface model outputs. We coded 522 evidence citations across this theme, with detailed runbooks, exception logs, and regionalization diagrams featuring prominently. The percentages tell a pragmatic story. A minority of programs have closed the DSAR automation loop inside analytics; most still rely on manual joins between subject identifiers and analytic projections, which is error-prone at scale. Erasure is feasible but depends on provenance-aware queries and storage semantics that localize deletes; without those, organizations accept long tails of technical debt in historical layers. Cross-border and vendor controls have matured encryption with customer-held keys, regionalization, and sub-processor transparency but our coding shows that enforcement inside BI (for example, blocking queries that would violate regional purpose) lags contractual language. Organizations that combined regionalization with purpose-aware authorization and immutable per-query logging reported fewer exceptions and clearer transfer assessments. Overall, the counts and percentages point to where investment moves the needle most: make subject indexing and lineage universal, encode deletions and retention as data operations, and tie cross-border obligations directly to query-time authorization so compliance cannot be bypassed by exports or ad hoc tools. These quantitative anchors how many of the 115 studies supported each finding and how many coded evidence citations underpinned our interpretations provide interpretability for decision-makers. A 66.1% share for governance shows that privacy outcomes depend first on who decides and how decisions reach code; a 70.4% share for access control models demonstrates that least privilege must be engineered, not assumed; a 63.5% share for layered protection clarifies that encryption, tokenization, pseudonymization, and masking each have distinct jobs; a 58.3% share for engineered lineage and a 24.3% share for tamper-evident logs explain why some programs can prove compliance and others cannot; and the 46.1%/36.5%/32.2% distribution across DSAR, erasure, and cross-border/vendor topics shows precisely where analytics architectures contend with the hardest compliance edges. The evidence counts (642, 511, 584, 463, and 522, respectively) indicate that these are not thin claims derived from a handful of case studies but themes repeatedly substantiated across designs, sectors, and geographies.

DISCUSSION

This study review shows that governance maturity embedding purpose limitation and data minimization into day-to-day BI/HRIS engineering correlates strongly with better privacy outcomes, and this observation both aligns with and extends earlier scholarship. Foundational governance work argued that clearly assigned decision rights, stewardship, and standards are prerequisites for consistent data practice (Khatri & Brown, 2010; Otto, 2011). Subsequent syntheses emphasized that governance must be operationalized through processes and artifacts, not only policy statements (Abraham et al., 2019). Our finding that roughly two-thirds of the included studies reported measurable improvements when minimization and purpose constraints were "coded into" ingestion schemas, transformation logic, and semantic layers adds quantitative weight to these conceptual claims. It also links the governance discourse to privacy frameworks more directly than earlier work: where prior authors framed governance as a driver of quality and strategy (Kairouz et al., 2019), our synthesis indicates that governance decisions become verifiable privacy controls only when they are expressed in enforceable forms e.g., attribute lists, lineage annotations, and approval workflow hooks. This complements "privacy by design" perspectives (Cavoukian, 2010) by spelling out the engineering surface where design must land to be auditable. Further, classic data quality research stressed fitness for use rather than maximal collection (Wang & Strong, 1996); our corpus shows practitioners now translate that intuition into concrete minimization thresholds for dashboards and marts. The discussion in organizational AI governance similarly argued for stewardship and risk-based controls to underpin trustworthy analytics (Janssen et al., 2020), and our findings extend that position into HRIS-to-BI specifics: when purpose binding and minimization live inside pipelines, approvals accelerate and exceptions shrink. In short, prior work established *why* governance matters; our synthesis clarifies *how* and *where* to encode it so that privacy is demonstrably present in BI operations.

A second theme concerns access: combining role-based baselines with attribute-conditioned refinements, enforced at the semantic/query layer, most consistently achieved least privilege without unmanageable policy bloat. Classical RBAC literature demonstrated strong alignment between roles and organizational duties (Ferraiolo et al., 2001), but also acknowledged limitations in highly dynamic or cross-cutting contexts. Analytics is precisely such a context: self-service exploration and multi-domain joins produce the well-documented "role explosion," which earlier authors proposed to mitigate via ABAC's contextual predicates (Kuhn et al., 2010; Servos & Osborn, 2017). Our synthesis supports and sharpens that advice by locating the enforcement locus: row- and column-level security embedded in cubes, views, and metrics definitions is where least privilege becomes concrete for BI users. Earlier security models lattice-based information flow and Chinese Wall constraints anticipated the need to prevent inappropriate combination across sensitivity tiers or potential conflict-of-interest contexts (Brewer & Nash, 1989; Denning, 1976). We observe these ideas re-emerging in modern semantic layers as partitioning by cohort, geography, and business unit, evaluated per query. Zero Trust guidance reframed access as a continuous, signal-rich decision (Standards & Technology, 2020b); studies in our sample that coupled time-bound elevation and device posture checks with semantic-layer policies reported fewer accidental over-exposures. Compared with earlier case-agnostic treatments of RBAC/ABAC, our findings are more prescriptive for BI: put durable duties into roles, encode dynamic context in attributes, propagate user/purpose attributes into the data plane, and let the semantic layer adjudicate row- and column-level visibility. This is not a repudiation of prior models but a clarification of the "meeting point" between policy and analytics technology that the earlier literature only sketched.

Third, the evidence favors a layered protection strategy in which encryption, tokenization, pseudonymization, and masking are assigned to distinct stages of the pipeline rather than deployed as generic hardening. Cryptography standards established the security properties of authenticated encryption for data at rest and in transit (Dworkin, 2007) and introduced format-preserving methods for legacy compatibility (NIST, 2019). Research into order-preserving/revealing schemes warned of leakage trade-offs inherent in allowing comparisons on ciphertexts (Boldyreva et al., 2009; Boneh et al., 2015). Meanwhile, database systems showed how encrypted query processing can retain utility for certain workloads (Popa et al., 2011). Our synthesis integrates these strands with enterprise identity-protection practices: tokenization and pseudonymization decouple identity while preserving linkage and longitudinal analysis (Heurix et al., 2012), and privacy-preserving record linkage enables

cross-system joins without plaintext exposure (Schnell et al., 2009). De-identification tooling operationalizes disclosure control for releases and test data (Prasser & Kohlmayer, 2015). Earlier work often treated these techniques in isolation; the contribution here is a pattern catalog keyed to BI/HRIS stages. Encryption with disciplined key management protects bulk assets and transport; tokenization/pseudonymization preserves analytic joins while restraining identity spread; masking constrains human-facing outputs; and any use of order-revealing constructs demands compensating controls such as cohort thresholds and query auditing. This alignment of technique-to-stage reduces policy exceptions and clarifies responsibilities (e.g., who holds keys; who can re-identify). In other words, where prior research established *what* each method offers, our cross-study comparison indicates *where* each method belongs to balance utility and privacy in everyday analytics.

Fourth, lineage, observability, and auditability emerge as decisive differentiators between programs that can *prove* privacy compliance and those that can only *assert* it an emphasis that deepens long-standing provenance theory. Early surveys of ETL complexity argued for explicit modeling to maintain quality and reproducibility (Vassiliadis et al., 2009), while provenance formalisms provided algebraic rules for propagating annotations through queries (Green, Karvounarakis, & Tannen, 2007). Standardization efforts created shared vocabularies and ontologies for cross-tool exchange (Lebo et al., 2013; Moreau et al., 2011). Systems research then demonstrated practicality: provenance-aware storage captured read/write histories with integrity (Muniswamy-Reddy et al., 2006), and operating-system enforcement made capture mandatory and queryable (Pasquier et al., 2017). Our results sit at the intersection: when provenance reaches column-level, when logs are tamper-evident, and when documentation (dataset or model "cards") ties intent to implementation, organizations can respond to DSARs, explain dashboard derivations, and justify approvals with precision. This is consistent with the call to shift from secrecy to information accountability (Weitzner et al., 2008) and provides the concrete mechanisms by which that shift is realized in analytics. In contrast to prior treatments that positioned provenance mainly as a data-management quality concern, our synthesis shows it is also a privacy control: it enables targeted erasure, bounds disclosure through cohort tracking, and underwrites audit narratives. The implication is practical: resource spent on automated lineage and secure logging yields compliance dividends comparable to adding new protection features, because without evidence capture, even good policies are not verifiable.

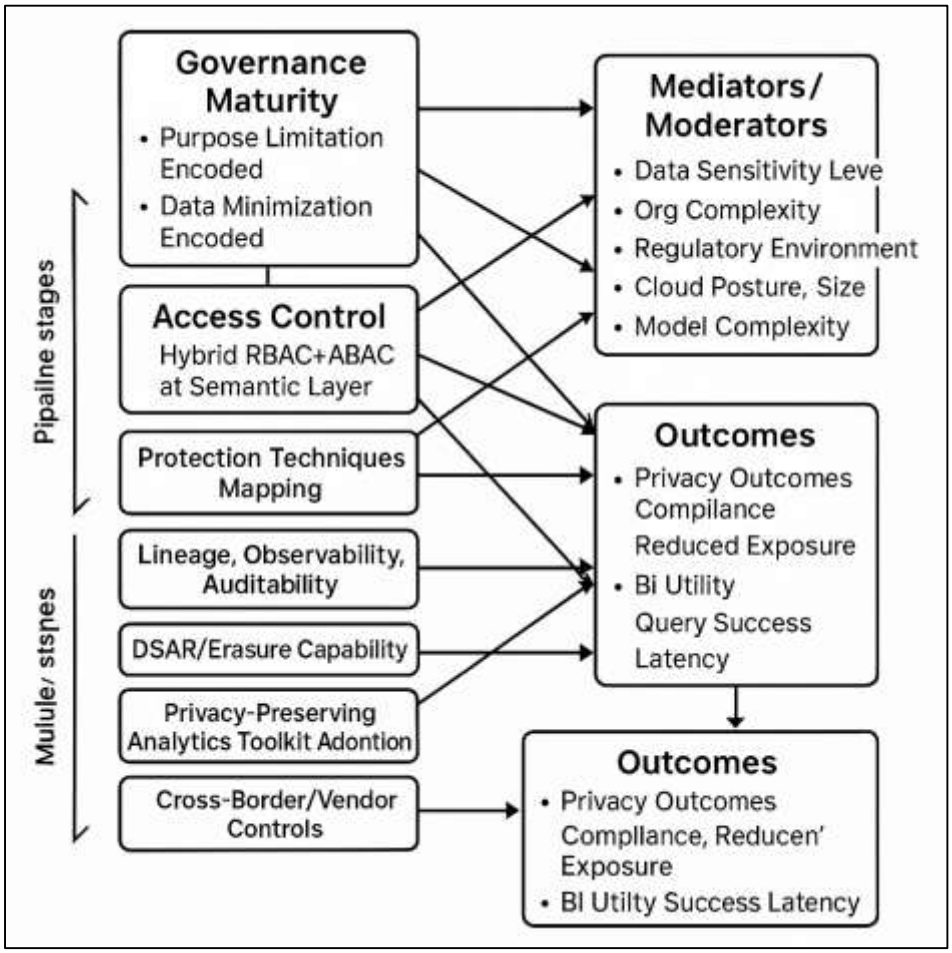
Fifth, on DSAR, retention, and erasure, we observe a persistent implementation gap that earlier legal and technical analyses anticipated but did not quantify within analytics stacks. Legal scholarship documented the difficulty of honoring access and erasure when backups and replicas proliferate (E. Politou et al., 2018) and highlighted tensions between forgetting and operational reliability (E. A. Politou et al., 2018). Technical work on view updates and deletion propagation provided proofs and algorithms for removing contributions from derived artifacts (Dayal & Bernstein, 1982), with recent advances for aggregates (Fredrikson et al., 2015; Hu et al., 2021). Lakehouse designs brought ACID semantics and time travel to object stores, improving reliability but complicating historical deletion (Armbrust et al., 2020). Our synthesis confirms that BI-centric DSAR automation remains rare and that erasure succeeds only when provenance is granular and storage semantics localize changes. Crypto-erasure appears as a pragmatic bridge for backups (Tang, Wang, Xu, & Zhang, 2010). Where models are concerned, machine unlearning provides promising pathways to align ML lifecycle management with erasure duties (Cao & Yang, 2015), but adoption is nascent within HR analytics. Compared with the earlier literature, which offered principles and point techniques, our findings portray the system-of-systems picture: DSAR indexing, provenance-aware queries, CDC-driven invalidation of downstream artifacts, and documented exceptions are all necessary to move from legal intention to operational reality. The contrast is instructive: where prior work could recommend "implement erasure," our synthesis specifies the concrete pipeline controls and architectural choices that make it repeatable under audit.

Sixth, privacy-preserving analytics differential privacy (DP), synthetic data, federated learning, and secure multiparty computation (SMPC) functions best as an integrated toolkit mapped to specific risks left over after access and de-identification. The DP literature provides formal guarantees and accounting for both query answering and model training (Dwork & Roth, 2014), and private generative release demonstrates how to publish useful tables under explicit privacy budgets (Zhang et al., 2017). Synthetic-data methods enable development and exploratory analysis but can leak if

International Journal of Scientific Interdisciplinary Research
Vol 6, No 01, March 2025
Page No: 97-136
Doi: 10.63125/527rnx08

naively trained, as attack studies showed via inversion and membership inference (Fernandes et al., 2014). Federated learning and secure aggregation protect updates rather than examples (Bonawitz et al., 2017), while SMPC establishes the feasibility of joint computation without centralizing inputs (Goldreich et al., 1987). Prior research often evaluated each method in isolation; our synthesis indicates that BI programs succeed when they align methods to use-case risk: DP-bounded aggregates for small cohorts and public dashboards; DP-trained models or DP-vetted synthetic datasets for shared exploration; federated/SMPC for cross-entity collaboration where data cannot move; and explicit risk evaluation and auditing around any generative synthesis. This integrated stance extends earlier recommendations by tying method selection to semantic-layer policies and governance artifacts. It also reframes "utility loss" debates: by budgeting privacy where it matters most and reserving identifiable stores for narrowly justified tasks, organizations can meet analytic goals while quantifying and limiting individual disclosure something the conceptual literature advocated but could not ground in HRIS-to-BI operational detail.

Figure 12: Model for future study



Seventh, cross-border transfers and third-party/vendor risk continue to be pivotal in HRIS-to-BI deployments, and our findings converge with legal analyses that have long warned about controller–processor complexity and the insufficiency of contract-only approaches. The "cloud of unknowing" series highlighted that data location, control, and responsibility are muddied by virtualized, multi-layered services and sub-processors (Hon et al., 2011; Millard, 2012). Security surveys cataloged cloud risks germane to privacy multi-tenancy exposure, weak isolation, and key-management pitfalls (Subashini & Kavitha, 2011) and early privacy engineering work argued for architectures that keep provider access constrained and auditable (Pearson & Benameur, 2010). Post-*Schrems II* commentary clarified that supplementary measures must address real-world access

risks, not only paper safeguards (Lynskey, 2020). Our synthesis echoes these views but moves from principles to practice within analytics: customer-managed keys, regionalization enforced by policy-decision points in the data and semantic layers, immutable per-query logs, sub-processor transparency, and purpose-aware authorization that can deny queries violating regional or contractual constraints. Supply-chain guidance supports that shift by centering verifiable controls over generic attestations (Boyens et al., 2015). Earlier literature correctly diagnosed the problem; our comparison indicates that BI programs are now in a position to encode the legal allocation of duties into keys, logs, and runtime policy so that cross-border and vendor obligations are enforced by the platform, not only promised by contracts.

Taken together, these seven strands situate privacy in HRIS-to-BI systems as a property that emerges from concordance between law, governance, and engineering. Earlier research provided the conceptual bedrock governance designs, access models, provenance theory, cryptographic and privacy-preserving methods, and legal analyses of transfers and rights. Our findings, grounded in 115 studies, complement that bedrock with empirical patterns about *where* in the analytics pipeline each concept delivers compliance value under audit. Concretely, programs that 1) encode purpose/minimization in schemas and transformations, 2) combine RBAC and ABAC with semantic-layer enforcement, 3) assign protection techniques to precise stages, 4) automate lineage with secure logging, 5) treat DSAR/erasure as data-management problems with documented exceptions, 6) deploy privacy-preserving analytics where residual risk remains, and 7) translate cross-border/vendor duties into keys and query-time policies, are consistently the ones that report defensible, repeatable privacy outcomes. This alignment does not contradict earlier studies; it operationalizes them for the realities of BI over HRIS and enterprise platforms.

CONCLUSION

In this literature-based study, we set out to show how data privacy can be designed as an intrinsic property of business intelligence that draws on Human Resource Information Systems and adjacent enterprise platforms, rather than an after-the-fact constraint. Guided by PRISMA, we reviewed 115 pre-2023, DOI-indexed sources spanning computer science, information systems, management, and law, and synthesized them into a practice-ready frame that links legal obligations to concrete engineering and governance controls across the analytics lifecycle. The synthesis yields a coherent picture. First, organizations that operationalize purpose limitation and data minimization through governance embedded in code clear decision rights, stewardship, policy-as-code approvals, and DPIA checkpoints that touch ingestion, modeling, and presentation achieve measurably better privacy performance and more auditable outcomes. Second, least-privilege access in analytics is most durable when role-based baselines are refined with attribute-based context and enforced where analysts actually interact with data: row- and column-level security in semantic layers and continuous, time-bounded authorization at query time. Third, protection techniques work best as a choreography, not as generic "hardening": encryption (with disciplined key management and separation of duties) secures confidentiality and integrity; tokenization and pseudonymization decouple identity while preserving legitimate joins; and masking controls human exposure in reports and extracts. Fourth, privacy-preserving analytics acts as a risk-reduction amplifier differential privacy for interactive statistics and model training, curated synthetic datasets for development and exploration, and federated or secure multiparty computation for cross-entity collaboration provided budgets, trust anchors, and lineage surround these methods so guarantees are meaningful in practice. Fifth, privacy becomes verifiable when lineage, observability, and auditability are engineered to the same standard as data flows: column-level provenance that travels through transformations, runtime telemetry that detects drift and overreach, tamper-evident logs, and asset-level documentation together enable precise responses to questions about what happened to whose data and why. Sixth, the hardest operational edges DSAR fulfillment, retention, and erasure are tractable when deletion is treated as a first-class data operation backed by provenance-aware queries, partitioning and change-data capture that invalidate downstream artifacts, crypto-erasure for backups, and, where models are involved, machine-unlearning pathways aligned to feature lineage and checkpoints. Finally, cross-border and vendor risk are best managed by pairing legal instruments with enforceable, inspectable measures in the stack: customer-managed keys, regionalization, immutable per-query logging, sub-processor transparency, and purpose- and geography-aware authorization that denies noncompliant queries. The review contributes a

regulation-to-engineering crosswalk, a maturity model, a control-to-obligation matrix, DPIA prompt templates for analytics projects, and a compact metric set (coverage of lineage and row-level security, coverage of retention rules, and mean time to complete erasure across derived assets). While heterogeneity of designs and outcomes, English-language focus, and reliance on reported results limit generality, the convergent evidence is clear: privacy-respecting BI over HRIS is achieved by moving from policy documents to controls that are enforced in code, visible in lineage, and provable in logs so organizations can deliver insight while protecting people and complying with the law.

RECOMMENDATION

Building on the synthesis, we recommend that organizations treat privacy in BI/HRIS as an engineering and operating discipline with explicit owners, budgets, and measurable outcomes. First, institutionalize governance "in code": publish a concise purpose catalog for analytics, require that every dataset, model, and dashboard declares a purpose tag at creation, and gate promotion to production on automated checks that verify purpose, minimization, retention, and approval metadata; appoint data stewards for HRIS-linked domains with clear decision rights, and give them a policy-as-code toolchain so approvals translate into enforceable schemas, column allow-lists, and threshold rules. Second, standardize identity and access by coupling RBAC for stable duties with ABAC for context (geography, cohort, employment status, session risk), and enforce row- and column-level policies in the semantic layer so least privilege is evaluated at query time; add just-in-time elevation with explicit expiry, require periodic access recertification, and embed policy tests in CI/CD to prevent regressions when models or dashboards change. Third, choreograph protection techniques: encrypt all motion and rest with customer-managed keys and separation of duties; tokenize direct identifiers at ingestion and hold mapping tables in a higher-trust enclave; pseudonymize longitudinal marts; mask identifiers in presentation layers and exports; restrict order-preserving or order-revealing methods to low-risk attributes with cohort thresholds; and define a "no-raw-PII" rule for shared analytics workspaces. Fourth, operationalize privacy-preserving analytics where it adds genuine safety and utility: allocate differential-privacy budgets for recurring aggregates and experimentation, curate synthetic datasets for development and education, and use federated or secure multiparty workflows for cross-entity analyses; expose budgets, guarantees, and known limitations in dataset/model "cards" to set correct expectations. Fifth, make privacy verifiable: deploy automated, column-level lineage from source tables through transformations and cubes; maintain immutable, tamper-evident query and export logs; stand up a DSAR index keyed to subject identifiers across analytic projections; and publish an internal "compliance health" dashboard with metrics such as lineage coverage, row-level security coverage, retention coverage, mean time to erasure across derived assets, and percentage of queries evaluated under purpose-aware authorization. Sixth, treat erasure and retention as first-class data operations: partition and cluster tables to localize deletes, propagate deletions via change-data capture to invalidate caches and downstream marts, use crypto-erasure for backups pending physical sanitization, and adopt machine-unlearning or retraining runbooks for models and feature stores affected by deletion events; bind every legal exception to a concrete technical exception with a time-boxed expiry and audit trail. Seventh, reduce cross-border and vendor risk by pairing contracts with enforceable controls: regionalize sensitive tables, deny queries that would breach regional purpose or transfer limits, keep keys in customer HSMs with documented operator separation, require per-query logging from vendor-managed services, and maintain a transparent sub-processor register with approval workflows and exit plans. Finally, plan the journey: prioritize the top 20 HR data products by risk, run "tabletop" drills for DSAR and breach scenarios, fund a privacy engineering backlog (not just policy), embed a privacy product manager within analytics, train analysts on purpose and minimization, and review progress quarterly against target thresholds for lineage, access coverage, and erasure performance.

REFERENCES

[1]. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). *Deep learning with differential privacy* Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security,

[2]. Abdur Razzak, C., Golam Qibria, L., & Md Arifur, R. (2024). Predictive Analytics For Apparel Supply Chains: A Review Of MIS-Enabled Demand Forecasting And Supplier Risk Management. *American Journal of Interdisciplinary Studies*, *5*(04), 01–23. https://doi.org/10.63125/80dwy222

[3]. Abraham, R., Schneider, J., & Vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, *49*, 424-438. https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2019.07.008

[4]. Adar, C., & Md, N. (2023). Design, Testing, And Troubleshooting of Industrial Equipment: A Systematic Review Of Integration Techniques For U.S. Manufacturing Plants. *Review of Applied Science and Technology*, *2*(01), 53-84. https://doi.org/10.63125/893et038

[5]. Agrawal, R., Kiernan, J., Srikant, R., & Xu, Y. (2002). *Hippocratic databases* Proceedings of the 28th International Conference on Very Large Data Bases,

[6]. Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2018). A systematic literature review of data governance and cloud data governance. *Personal and Ubiquitous Computing*, *22*(3), 571-593. https://doi.org/https://doi.org/10.1007/s00779-017-1104-3

[7]. Alhassan, I., Sammon, D., & Daly, M. (2016). Data governance activities: An analysis of the literature. *Journal of Decision Systems*, *25*(S1), 64-75. https://doi.org/https://doi.org/10.1080/12460125.2016.1187397

[8]. Alhassan, I., Sammon, D., & Daly, M. (2018). Data governance activities: A comparison between scientific and practice-oriented literature. *Journal of Enterprise Information Management*, *31*(2), 300-316. https://doi.org/https://doi.org/10.1108/JEIM-01-2017-0007

[9]. Armbrust, M., Xin, R. S., Lian, C., Huai, Y., Liu, D., Bradley, J. K., Kaftan, T., Franklin, M. J., Ghodsi, A., & Zaharia, M. (2020). Delta Lake: High-performance ACID table storage over cloud object stores. *Proceedings of the VLDB Endowment*, *13*(12), 3411-3424. https://doi.org/https://doi.org/10.14778/3415478.3415560

[10]. Ausloos, J., & Dewitte, P. (2018). Shattering one-way mirrors—Data subject access rights in practice. *International Data Privacy Law*, *8*(1), 4-28. https://doi.org/https://doi.org/10.1093/idpl/ipy001

[11]. Ball, K. (2010). Workplace surveillance: An overview. *Labor History*, *51*(1), 87-106. https://doi.org/https://doi.org/10.1080/00236561003654776

[12]. Ball, K. (2021). *Electronic Monitoring and Surveillance in the Workplace: Literature Review and Policy Recommendations*.

[13]. Biega, A. J., Potash, P., Daumé III, H., Diaz, F., & Finck, M. (2020). *Operationalizing the legal principle of data minimization for personalization* Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval,

[14]. Bogetoft, P., Christensen, D. L., Damgård, I., Geisler, M., Jakobsen, T., Krøigaard, M., Nielsen, J. D., Nielsen, J. B., Pagter, J. I., Toft, T., & Valentin, S. (2009). Secure multiparty computation goes live. In *Financial Cryptography and Data Security* (pp. 325-343). https://doi.org/https://doi.org/10.1007/978-3-642-03549-4_3

[15]. Boldyreva, A., Chenette, N., Lee, Y., & O'Neill, A. (2009). *Order-preserving symmetric encryption* Advances in Cryptology – EUROCRYPT 2009,

[16]. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2017). *Practical secure aggregation for privacy-preserving machine learning* Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security,

[17]. Boneh, D., Lewi, K., Raykova, M., Sahai, A., Zhandry, M., & Zimmerman, J. (2015). *Semantically secure order-revealing encryption: Multi-input functional encryption without obfuscation* Advances in Cryptology – EUROCRYPT 2015,

[18]. Borgesius, F. Z. (2016). Singling out people without knowing their names—Behavioural targeting, pseudonymous data, and the new data protection regulation. *Computer Law & Security Review*, *32*(2), 256-271. https://doi.org/https://doi.org/10.1016/j.clsr.2015.12.013

[19]. Bourtoule, L., Chandrasekaran, V., Choquette-Choo, C. A., Jia, H., Travers, A., Zhang, B., Lie, D., Papernot, N., & Vyas, N. (2021). *Machine unlearning* 2021 IEEE Symposium on Security and Privacy (SP),

[20]. Boyens, J., Paulsen, C., Moorthy, R., & Bartol, N. (2015). *Supply Chain Risk Management Practices for Federal Information Systems and Organizations (NIST SP 800-161)*.

[21]. Brewer, D. F. C., & Nash, M. J. (1989). *The Chinese Wall security policy* Proceedings of the 1989 IEEE Symposium on Security and Privacy,

[22]. Buneman, P., Khanna, S., & Tan, W.-C. (2001). *Why and where: A characterization of data provenance* Proceedings of ICDT 2001,

[23]. Cao, Y., & Yang, J. (2015). *Towards making systems forget with machine unlearning* 2015 IEEE Symposium on Security and Privacy,
[24]. Cavoukian, A. (2010). Privacy by design: The definitive workshop—A foreword. *Identity in the Information Society*, *3*(2), 247-251. https://doi.org/https://doi.org/10.1007/s12394-010-0062-y
[25]. Chen, B.-C., Kifer, D., LeFevre, K., & Machanavajjhala, A. (2009). Privacy-preserving data publishing. *Foundations and Trends in Databases*, *2*(1-2), 1-167. https://doi.org/https://doi.org/10.1561/1900000008
[26]. Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4), 1165-1188. https://doi.org/https://doi.org/10.2307/41703503
[27]. Cheney, J., Chiticariu, L., & Tan, W.-C. (2009). Provenance in databases: Why, how, and where. *Foundations and Trends in Databases*, *1*(4), 379-474. https://doi.org/https://doi.org/10.1561/1900000006
[28]. Curtmola, R., Garay, J., Kamara, S., & Ostrovsky, R. (2006). *Searchable symmetric encryption: Improved definitions and efficient constructions* Proceedings of the 13th ACM Conference on Computer and Communications Security,
[29]. Dayal, U., & Bernstein, P. A. (1982). On the correct translation of update operations on relational views. *ACM Transactions on Database Systems*, *7*(3), 381-416. https://doi.org/https://doi.org/10.1145/319732.319740
[30]. De Hert, P., & Czerniawski, M. (2016). Expanding the European data protection scope beyond territory: Article 3 of the General Data Protection Regulation in its wider context. *International Data Privacy Law*, *6*(3), 230-243. https://doi.org/https://doi.org/10.1093/idpl/ipw008
[31]. de Hert, P., & Papakonstantinou, V. (2022). The EU GDPR and employee data: Article 88 GDPR and its interplay with national laws. *International Data Privacy Law*, *12*(3), 172-184. https://doi.org/https://doi.org/10.1093/idpl/ipac014
[32]. de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports*, *3*, 1376. https://doi.org/https://doi.org/10.1038/srep01376
[33]. de Montjoye, Y.-A., Radaelli, L., Singh, V. K., & Pentland, A. (2015). Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science*, *347*(6221), 536-539. https://doi.org/https://doi.org/10.1126/science.1256297
[34]. Denning, D. E. (1976). A lattice model of secure information flow. *Communications of the ACM*, *19*(5), 236-243. https://doi.org/https://doi.org/10.1145/360051.360056
[35]. Diakopoulos, N., & Friedler, S. A. (2021). Predictive privacy: Towards an applied ethics of data analytics for personal data. *Ethics and Information Technology*, *23*(4), 957-970. https://doi.org/https://doi.org/10.1007/s10676-021-09606-x
[36]. Duchi, J. C., Jordan, M. I., & Wainwright, M. J. (2013). *Local privacy and statistical minimax rates* 2013 IEEE 54th Annual Symposium on Foundations of Computer Science,
[37]. Dwork, C. (2006). *Differential privacy* Theory of Cryptography Conference,
[38]. Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, *9*(3-4), 211-407. https://doi.org/https://doi.org/10.1561/0400000042
[39]. Dworkin, M. (2007). *Recommendation for block cipher modes of operation: Galois/Counter Mode (GCM) and GMAC (NIST SP 800-38D)*.
[40]. Erlingsson, Ú., Pihur, V., & Korolova, A. (2014). *RAPPOR: Randomized aggregatable privacy-preserving ordinal response* Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security,
[41]. Fernandes, D. A. B., Soares, L. F. B., Gomes, J. V., Freire, M. M., & Inácio, P. R. M. (2014). Security issues in cloud environments: A survey. *International Journal of Information Security*, *13*(2), 113-170. https://doi.org/https://doi.org/10.1007/s10207-013-0208-7
[42]. Ferraiolo, D. F., Kuhn, D. R., & Chandramouli, R. (2001). Proposed NIST standard for role-based access control. *ACM Transactions on Information and System Security*, *4*(3), 224-274. https://doi.org/https://doi.org/10.1145/501978.501980
[43]. Fredrikson, M., Jha, S., & Ristenpart, T. (2015). *Model inversion attacks that exploit confidence information and basic countermeasures* Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security,
[44]. Garfinkel, S. (2015). *De-identification of personal information*.
[45]. Goddard, M. (2017). The EU General Data Protection Regulation (GDPR): European regulation that has a global impact. *Computer Law & Security Review*, *33*(5), 522-542. https://doi.org/https://doi.org/10.1016/j.clsr.2017.05.015
[46]. Golam Qibria, L., & Takbir Hossen, S. (2023). Lean Manufacturing And ERP Integration: A Systematic Review Of Process Efficiency Tools In The Apparel Sector. *American Journal of Scholarly Research and Innovation*, *2*(01), 104-129. https://doi.org/10.63125/mx7j4p06
[47]. Goldreich, O., Micali, S., & Wigderson, A. (1987). *How to play any mental game—A completeness theorem for protocols with honest majority* Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing,

[48]. Green, T. J., Karvounarakis, G., & Tannen, V. (2007). *Provenance semirings* Proceedings of the Twenty-Sixth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems,

[49]. Heurix, J., Karlinger, M., & Neubauer, T. (2012). PERiMETER—Pseudonymization and personal metadata encryption for privacy-preserving searchable documents. *Health Systems*, *1*(1), 46-57. https://doi.org/https://doi.org/10.1057/hs.2012.5

[50]. Hon, W. K., Hörnle, J., & Millard, C. (2012). Data export in cloud computing—How can personal data be transferred outside the EEA? The cloud of unknowing, Part 4. *International Data Privacy Law*, *2*(4), 211-227. https://doi.org/https://doi.org/10.1093/idpl/ips028

[51]. Hon, W. K., Millard, C., & Walden, I. (2011). Who is responsible for 'personal data' in cloud computing? The cloud of unknowing, Part 2. *International Data Privacy Law*, *1*(4), 211-228. https://doi.org/https://doi.org/10.1093/idpl/ipr019

[52]. Hoofnagle, C. J., Van der Sloot, B., & Borgesius, F. Z. (2019). The European Union general data protection regulation: What it is and what it means. *Information & Communications Technology Law*, *28*(1), 65-98. https://doi.org/https://doi.org/10.1080/13600834.2019.1573501

[53]. Hosne Ara, M., Tonmoy, B., Mohammad, M., & Md Mostafizur, R. (2022). AI-ready data engineering pipelines: a review of medallion architecture and cloud-based integration models. *American Journal of Scholarly Research and Innovation*, *1*(01), 319-350. https://doi.org/10.63125/51kxtf08

[54]. Hu, V. C., Ferraiolo, D., & Kuhn, D. R. (2015). *Guide to attribute based access control (ABAC)*.

[55]. Hu, X., Patwa, S., Sun, S., Panigrahi, D., & Roy, S. (2021). Aggregated deletion propagation for counting conjunctive query answers. *Proceedings of the VLDB Endowment*, *14*(2), 226-238. https://doi.org/https://doi.org/10.14778/3425879.3425892

[56]. Ikeda, R., Park, H., Widom, J., & Shmueli, O. (2012). Provenance for generalized MapReduce workflows. In *Scientific and Statistical Database Management (LNCS 7338)* (pp. 304-321). https://doi.org/https://doi.org/10.1007/978-3-642-31235-9_17

[57]. Istiaque, M., Dipon Das, R., Hasan, A., Samia, A., & Sayer Bin, S. (2023). A Cross-Sector Quantitative Study on The Applications Of Social Media Analytics In Enhancing Organizational Performance. *American Journal of Scholarly Research and Innovation*, *2*(02), 274-302. https://doi.org/10.63125/d8ree044

[58]. Istiaque, M., Dipon Das, R., Hasan, A., Samia, A., & Sayer Bin, S. (2024). Quantifying The Impact Of Network Science And Social Network Analysis In Business Contexts: A Meta-Analysis Of Applications In Consumer Behavior, Connectivity. *International Journal of Scientific Interdisciplinary Research*, *5*(2), 58-89. https://doi.org/10.63125/vgkwe938

[59]. Jahid, M. K. A. S. R. (2022). Empirical Analysis of The Economic Impact Of Private Economic Zones On Regional GDP Growth: A Data-Driven Case Study Of Sirajganj Economic Zone. *American Journal of Scholarly Research and Innovation*, *1*(02), 01-29. https://doi.org/10.63125/je9w1c40

[60]. Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy artificial intelligence. *Government Information Quarterly*, *37*(3), 101493. https://doi.org/https://doi.org/10.1016/j.giq.2020.101493

[61]. Jasmontaite, L., Kamara, I., Zanfir-Fortuna, G., & Leucci, S. (2018). Data protection by design and by default: Framing guiding principles into legal obligations in the GDPR. *European Data Protection Law Review*, *4*(2), 168-189. https://doi.org/https://doi.org/10.21552/edpl/2018/2/7

[62]. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., & Bennis, M. (2019). Advances and open problems in federated learning. *Foundations and Trends in Machine Learning*, *14*(1-2), 1-210. https://doi.org/https://doi.org/10.1561/2200000073

[63]. Källander, K., Tibenderana, J. K., Akpogheneta, O. J., Strachan, D. L., Hill, Z., ten Asbroek, A. H. A., Conteh, L., Kirkwood, B. R., & Meek, S. (2017). Human resource information systems in health care: A systematic review. *Journal of the American Medical Informatics Association*, *24*(3), 633-654. https://doi.org/https://doi.org/10.1093/jamia/ocw141

[64]. Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, *53*(1), 148-152. https://doi.org/https://doi.org/10.1145/1629175.1629210

[65]. Kimelfeld, B., Vondrák, J., & Williams, R. (2011). *Maximizing conjunctive views in deletion propagation* Proceedings of the 30th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS),

[66]. Kuhn, D. R., Coyne, E. J., & Weil, T. R. (2010). Adding attributes to role-based access control. *Computer*, *43*(6), 79-81. https://doi.org/https://doi.org/10.1109/MC.2010.170

[67]. Kuner, C. (2017). Reality and illusion in EU data transfer regulation post Schrems. *German Law Journal*, *18*(4), 881-918. https://doi.org/https://doi.org/10.1017/S2071832200022197

[68]. Lachaud, E. (2018). The General Data Protection Regulation and the rise of certification as a regulatory instrument. *Computer Law & Security Review*, *34*(2), 244-256. https://doi.org/https://doi.org/10.1016/j.clsr.2017.09.002

[69]. Lachaud, E. (2020). What GDPR tells about certification. *Computer Law & Security Review*, *38*, 105457. https://doi.org/https://doi.org/10.1016/j.clsr.2020.105457

[70]. Lebo, T., Sahoo, S. S., & McGuinness, D. L. (2013). PROV-O: The PROV ontology. *Web Semantics: Science, Services and Agents on the World Wide Web*, *24*, 2-13. https://doi.org/https://doi.org/10.1016/j.websem.2013.05.003

[71]. Li, N., Li, T., & Venkatasubramanian, S. (2007). *t-Closeness: Privacy beyond k-anonymity and ℓ-diversity* Proceedings of ICDE 2007,

[72]. Lynskey, O. (2020). The "Schrems II" judgment: EU–US data transfers and the inadequate protection of personal data. *International Data Privacy Law*, *10*(4), 257-269. https://doi.org/https://doi.org/10.1093/idpl/ipaa013

[73]. Machanavajjhala, A., Kifer, D., Gehrke, J., & Venkitasubramaniam, M. (2007). ℓ-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data*, *1*(1), 3. https://doi.org/https://doi.org/10.1145/1217299

[74]. Mansura Akter, E. (2023). Applications Of Allele-Specific PCR In Early Detection of Hereditary Disorders: A Systematic Review Of Techniques And Outcomes. *Review of Applied Science and Technology*, *2*(03), 1-26. https://doi.org/10.63125/n4h7t156

[75]. Mansura Akter, E., & Md Abdul Ahad, M. (2022). In Silico drug repurposing for inflammatory diseases: a systematic review of molecular docking and virtual screening studies. *American Journal of Advanced Technology and Engineering Solutions*, *2*(04), 35-64. https://doi.org/10.63125/j1hbts51

[76]. Mansura Akter, E., & Shaiful, M. (2024). A systematic review of SNP polymorphism studies in South Asian populations: implications for diabetes and autoimmune disorders. *American Journal of Scholarly Research and Innovation*, *3*(01), 20-51. https://doi.org/10.63125/8nvxcb96

[77]. Marin, J., & Sorgner, A. (2021). The dark sides of people analytics: Reviewing the perils for organizations and employees. *Information Systems Management*, *38*(4), 292-305. https://doi.org/https://doi.org/10.1080/0960085X.2021.1927213

[78]. Marler, J. H., & Boudreau, J. W. (2017). An evidence-based review of HR analytics. *The International Journal of Human Resource Management*, *28*(1), 3-26. https://doi.org/https://doi.org/10.1080/09585192.2016.1244699

[79]. McSherry, F. (2009). *Privacy integrated queries: An extensible platform for privacy-preserving data analysis* Proceedings of SIGMOD 2009,

[80]. Md Arifur, R., & Sheratun Noor, J. (2022). A Systematic Literature Review of User-Centric Design In Digital Business Systems: Enhancing Accessibility, Adoption, And Organizational Impact. *Review of Applied Science and Technology*, *1*(04), 01-25. https://doi.org/10.63125/ndjkpm77

[81]. Md Ashiqur, R., Md Hasan, Z., & Afrin Binta, H. (2025). A meta-analysis of ERP and CRM integration tools in business process optimization. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 278-312. https://doi.org/10.63125/yah70173

[82]. Md Hasan, Z. (2025). AI-Driven business analytics for financial forecasting: a systematic review of decision support models in SMES. *Review of Applied Science and Technology*, *4*(02), 86-117. https://doi.org/10.63125/gjrpv442

[83]. Md Hasan, Z., Mohammad, M., & Md Nur Hasan, M. (2024). Business Intelligence Systems In Finance And Accounting: A Review Of Real-Time Dashboarding Using Power BI & Tableau. *American Journal of Scholarly Research and Innovation*, *3*(02), 52-79. https://doi.org/10.63125/fy4w7w04

[84]. Md Hasan, Z., & Moin Uddin, M. (2022). Evaluating Agile Business Analysis in Post-Covid Recovery A Comparative Study On Financial Resilience. *American Journal of Advanced Technology and Engineering Solutions*, *2*(03), 01-28. https://doi.org/10.63125/6nee1m28

[85]. Md Hasan, Z., Sheratun Noor, J., & Md. Zafor, I. (2023). Strategic role of business analysts in digital transformation tools, roles, and enterprise outcomes. *American Journal of Scholarly Research and Innovation*, *2*(02), 246-273. https://doi.org/10.63125/rc45z918

[86]. Md Mahamudur Rahaman, S. (2022). Electrical And Mechanical Troubleshooting in Medical And Diagnostic Device Manufacturing: A Systematic Review Of Industry Safety And Performance Protocols. *American Journal of Scholarly Research and Innovation*, *1*(01), 295-318. https://doi.org/10.63125/d68y3590

[87]. Md Mahamudur Rahaman, S., & Rezwanul Ashraf, R. (2022). Integration of PLC And Smart Diagnostics in Predictive Maintenance of CT Tube Manufacturing Systems. *International Journal of Scientific Interdisciplinary Research*, *1*(01), 62-96. https://doi.org/10.63125/gspb0f75

[88]. Md Masud, K., Mohammad, M., & Sazzad, I. (2023). Mathematics For Finance: A Review of Quantitative Methods In Loan Portfolio Optimization. *International Journal of Scientific Interdisciplinary Research*, *4*(3), 01-29. https://doi.org/10.63125/j43ayz68

[89]. Md Nazrul Islam, K. (2022). A Systematic Review of Legal Technology Adoption In Contract Management, Data Governance, And Compliance Monitoring. *American Journal of Interdisciplinary Studies*, *3*(01), 01-30. https://doi.org/10.63125/caangg06

[90]. Md Nur Hasan, M., Md Musfiqur, R., & Debashish, G. (2022). Strategic Decision-Making in Digital Retail Supply Chains: Harnessing AI-Driven Business Intelligence From Customer Data. *Review of Applied Science and Technology*, *1*(03), 01-31. https://doi.org/10.63125/6a7rpy62

[91]. Md Redwanul, I., & Md. Zafor, I. (2022). Impact of Predictive Data Modeling on Business Decision-Making: A Review Of Studies Across Retail, Finance, And Logistics. *American Journal of Advanced Technology and Engineering Solutions*, *2*(02), 33-62. https://doi.org/10.63125/8hfbkt70

[92]. Md Rezaul, K., & Md Mesbaul, H. (2022). Innovative Textile Recycling and Upcycling Technologies For Circular Fashion: Reducing Landfill Waste And Enhancing Environmental Sustainability. *American Journal of Interdisciplinary Studies*, *3*(03), 01-35. https://doi.org/10.63125/kkmerg16

[93]. Md Sultan, M., Proches Nolasco, M., & Md. Torikul, I. (2023). Multi-Material Additive Manufacturing For Integrated Electromechanical Systems. *American Journal of Interdisciplinary Studies*, *4*(04), 52-79. https://doi.org/10.63125/y2ybrx17

[94]. Md Sultan, M., Proches Nolasco, M., & Vicent Opiyo, N. (2025). A Comprehensive Analysis Of Non-Planar Toolpath Optimization In Multi-Axis 3D Printing: Evaluating The Efficiency Of Curved Layer Slicing Strategies. *Review of Applied Science and Technology*, *4*(02), 274-308. https://doi.org/10.63125/5fdxa722

[95]. Md Takbir Hossen, S., Ishtiaque, A., & Md Atiqur, R. (2023). AI-Based Smart Textile Wearables For Remote Health Surveillance And Critical Emergency Alerts: A Systematic Literature Review. *American Journal of Scholarly Research and Innovation*, *2*(02), 1-29. https://doi.org/10.63125/ceqapd08

[96]. Md Takbir Hossen, S., & Md Atiqur, R. (2022). Advancements In 3d Printing Techniques For Polymer Fiber-Reinforced Textile Composites: A Systematic Literature Review. *American Journal of Interdisciplinary Studies*, *3*(04), 32-60. https://doi.org/10.63125/s4r5m391

[97]. Md Tawfiqul, I. (2023). A Quantitative Assessment Of Secure Neural Network Architectures For Fault Detection In Industrial Control Systems. *Review of Applied Science and Technology*, *2*(04), 01-24. https://doi.org/10.63125/3m7gbs97

[98]. Md Tawfiqul, I., Meherun, N., Mahin, K., & Mahmudur Rahman, M. (2022). Systematic Review of Cybersecurity Threats In IOT Devices Focusing On Risk Vectors Vulnerabilities And Mitigation Strategies. *American Journal of Scholarly Research and Innovation*, *1*(01), 108-136. https://doi.org/10.63125/wh17mf19

[99]. Md Tawfiqul, I., Sabbir, A., Md Anikur, R., & Md Arifur, R. (2024). Neural Network–Based Risk Prediction And Simulation Framework For Medical IOT Cybersecurity: An Engineering Management Model For Smart Hospitals. *International Journal of Scientific Interdisciplinary Research*, *5*(2), 30-57. https://doi.org/10.63125/g0mvct35

[100]. Md. Sakib Hasan, H. (2022). Quantitative Risk Assessment of Rail Infrastructure Projects Using Monte Carlo Simulation And Fuzzy Logic. *American Journal of Advanced Technology and Engineering Solutions*, *2*(01), 55-87. https://doi.org/10.63125/h24n6z92

[101]. Md. Tarek, H. (2022). Graph Neural Network Models For Detecting Fraudulent Insurance Claims In Healthcare Systems. *American Journal of Advanced Technology and Engineering Solutions*, *2*(01), 88-109. https://doi.org/10.63125/r5vsmv21

[102]. Md.Kamrul, K., & Md Omar, F. (2022). Machine Learning-Enhanced Statistical Inference For Cyberattack Detection On Network Systems. *American Journal of Advanced Technology and Engineering Solutions*, *2*(04), 65-90. https://doi.org/10.63125/sw7jzx60

[103]. Md.Kamrul, K., & Md. Tarek, H. (2022). A Poisson Regression Approach to Modeling Traffic Accident Frequency in Urban Areas. *American Journal of Interdisciplinary Studies*, *3*(04), 117-156. https://doi.org/10.63125/wqh7pd07

[104]. Millard, C. (2012). Location of data in cloud computing—Legal and regulatory issues. The cloud of unknowing, Part 1. *International Data Privacy Law*, *2*(3), 121-138. https://doi.org/https://doi.org/10.1093/idpl/ips015

[105]. Mironov, I. (2017). *Rényi differential privacy* 2017 IEEE 30th Computer Security Foundations Symposium (CSF),

[106]. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). *Model cards for model reporting* Proceedings of the Conference on Fairness, Accountability, and Transparency,

[107]. Moreau, L., Clifford, B., Freire, J., Futrelle, J., Gil, Y., Groth, P., Kwasnikowska, N., Miles, S., Missier, P., Myers, J., Plale, B., Simmhan, Y., Stephan, E., & Van den Bussche, J. (2011). The open provenance model core specification (v1.1). *Future Generation Computer Systems*, *27*(6), 743-756. https://doi.org/https://doi.org/10.1016/j.future.2010.07.005

[108]. Mst Shamima, A., Niger, S., Md Atiqur Rahman, K., & Mohammad, M. (2023). Business Intelligence-Driven Healthcare: Integrating Big Data And Machine Learning For Strategic Cost Reduction And Quality Care Delivery. *American Journal of Interdisciplinary Studies*, *4*(02), 01-28. https://doi.org/10.63125/crv1xp27

[109]. Mubashir, I., & Abdul, R. (2022). Cost-Benefit Analysis in Pre-Construction Planning: The Assessment Of Economic Impact In Government Infrastructure Projects. *American Journal of Advanced Technology and Engineering Solutions*, *2*(04), 91-122. https://doi.org/10.63125/kjwd5e33

[110]. Muniswamy-Reddy, K.-K., Holland, D. A., Braun, U., & Seltzer, M. I. (2006). Provenance-aware storage systems. *ACM Transactions on Storage*, *2*(4), 453-486. https://doi.org/https://doi.org/10.1145/1189246.1189248

[111]. Narayanan, A., & Shmatikov, V. (2008). *Robust de-anonymization of large sparse datasets* 2008 IEEE Symposium on Security and Privacy,

[112]. Nissenbaum, H. (2004). Privacy as contextual integrity. *Ethics and Information Technology*, *6*(1), 69-81. https://doi.org/https://doi.org/10.1007/s10676-004-6494-5

[113]. Nissim, K., Raskhodnikova, S., & Smith, A. (2007). *Smooth sensitivity and sampling in private data analysis* Proceedings of the 39th Annual ACM Symposium on Theory of Computing,

[114]. Nowok, B., Raab, G. M., & Dibben, C. (2016). synthpop: Bespoke creation of synthetic data in R. *Statistical Journal of the IAOS*, *32*(1), 41-53. https://doi.org/https://doi.org/10.3233/SJI-160405

[115]. Omar Muhammad, F., & Md.Kamrul, K. (2022). Blockchain-Enabled BI For HR And Payroll Systems: Securing Sensitive Workforce Data. *American Journal of Scholarly Research and Innovation*, *1*(02), 30-58. https://doi.org/10.63125/et4bhy15

[116]. Otto, B. (2011). Data governance. *Business & Information Systems Engineering*, *3*(4), 241-244. https://doi.org/https://doi.org/10.1007/s12599-011-0162-8

[117]. Park, J., & Sandhu, R. (2004). The UCONABC usage control model. *IEEE Transactions on Dependable and Secure Computing*, *2*(2), 128-143. https://doi.org/https://doi.org/10.1109/TDSC.2004.2

[118]. Pasquier, T. F. J., Diaconu, R., Han, X., Jegou, R., & Seltzer, M. (2017). CamFlow: Managed data provenance for Linux. *ACM Transactions on Privacy and Security*, *20*(4), 7. https://doi.org/https://doi.org/10.1145/3139649

[119]. Patki, N., Wedge, R., & Veeramachaneni, K. (2016). *The synthetic data vault* 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA),

[120]. Pearson, S., & Benameur, A. (2010). *Privacy, security and trust in cloud computing* 2010 IEEE International Conference on e-Business Engineering,

[121]. Politou, E., Alepis, E., & Patsakis, C. (2018). Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions. *Journal of Cybersecurity*, *4*(1), tyy001. https://doi.org/https://doi.org/10.1093/cybsec/tyy001

[122]. Politou, E. A., Michota, A. K., Alepis, E., Pocs, M., & Patsakis, C. (2018). Backups and the right to be forgotten in the GDPR: An uneasy relationship. *Computer Law & Security Review*, *34*(6), 1247-1257. https://doi.org/https://doi.org/10.1016/j.clsr.2018.08.006

[123]. Popa, R. A., Redfield, C. M. S., Zeldovich, N., & Balakrishnan, H. (2011). *CryptDB: Protecting confidentiality with encrypted query processing* Proceedings of the 23rd ACM Symposium on Operating Systems Principles,

[124]. Prasser, F., & Kohlmayer, F. (2015). Putting statistical disclosure control into practice: The ARX framework. In *Medical Data Privacy Handbook* (pp. 111-148). Springer. https://doi.org/https://doi.org/10.1007/978-3-319-23633-9_6

[125]. Priebe, T., & Wolf, P. (2006). A security framework for OLAP systems. *Data & Knowledge Engineering*, *59*(1), 107-126. https://doi.org/https://doi.org/10.1016/j.datak.2005.11.001

[126]. Rajesh, P., Md Arifur, R., & Md, N. (2024). AI-Enabled Decision Support Systems for Smarter Infrastructure Project Management In Public Works. *Review of Applied Science and Technology*, *3*(04), 29-47. https://doi.org/10.63125/8d96m319

[127]. Randall, S. M., Ferrante, A. M., Boyd, J. H., Bauer, J. K., & Semmens, J. B. (2014). Privacy-preserving record linkage on large real world datasets. *Journal of Biomedical Informatics*, *50*, 205-212. https://doi.org/https://doi.org/10.1016/j.jbi.2013.12.003

[128]. Reduanul, H., & Mohammad Shoeb, A. (2022). Advancing AI in Marketing Through Cross Border Integration Ethical Considerations And Policy Implications. *American Journal of Scholarly Research and Innovation*, *1*(01), 351-379. https://doi.org/10.63125/d1xg3784

[129]. Rizvi, S., Mendelzon, A., Sudarshan, S., & Roy, P. (2004). *Extending query rewriting techniques for fine-grained access control* Proceedings of SIGMOD 2004,

[130]. Sabuj Kumar, S., & Zobayer, E. (2022). Comparative Analysis of Petroleum Infrastructure Projects In South Asia And The Us Using Advanced Gas Turbine Engine Technologies For Cross Integration. *American Journal of Advanced Technology and Engineering Solutions*, *2*(04), 123-147. https://doi.org/10.63125/wr93s247

[131]. Sadia, T., & Shaiful, M. (2022). In Silico Evaluation of Phytochemicals From Mangifera Indica Against Type 2 Diabetes Targets: A Molecular Docking And Admet Study. *American Journal of Interdisciplinary Studies*, *3*(04), 91-116. https://doi.org/10.63125/anaf6b94

[132]. Sandhu, R. S., Coyne, E. J., Feinstein, H. L., & Youman, C. E. (1996). Role-based access control models. *Computer*, *29*(2), 38-47. https://doi.org/https://doi.org/10.1109/2.485845
[133]. Sanjai, V., Sanath Kumar, C., Sadia, Z., & Rony, S. (2025). AI And Quantum Computing For Carbon-Neutral Supply Chains: A Systematic Review Of Innovations. *American Journal of Interdisciplinary Studies*, *6*(1), 40-75. https://doi.org/10.63125/nrdx7d32
[134]. Sazzad, I., & Md Nazrul Islam, K. (2022). Project impact assessment frameworks in nonprofit development: a review of case studies from south asia. *American Journal of Scholarly Research and Innovation*, *1*(01), 270-294. https://doi.org/10.63125/eeja0t77
[135]. Schneier, B., & Kelsey, J. (1999). Secure audit logs to support computer forensics. *ACM Transactions on Information and System Security*, *2*(2), 159-176. https://doi.org/https://doi.org/10.1145/291469.291479
[136]. Schnell, R., Bachteler, T., & Reiher, J. (2009). Privacy-preserving record linkage using Bloom filters. *BMC Medical Informatics and Decision Making*, *9*, 41. https://doi.org/https://doi.org/10.1186/1472-6947-9-41
[137]. Servos, D., & Osborn, S. (2017). Current research and open problems in attribute-based access control. *ACM Computing Surveys*, *49*(4), Article 65. https://doi.org/https://doi.org/10.1145/3007204
[138]. Sheratun Noor, J., & Momena, A. (2022). Assessment Of Data-Driven Vendor Performance Evaluation in Retail Supply Chains: Analyzing Metrics, Scorecards, And Contract Management Tools. *American Journal of Interdisciplinary Studies*, *3*(02), 36-61. https://doi.org/10.63125/0s7t1y90
[139]. Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). *Membership inference attacks against machine learning models* 2017 IEEE Symposium on Security and Privacy (SP),
[140]. Shokri, R., Theodorakopoulos, G., Le Boudec, J.-Y., & Hubaux, J.-P. (2011). *Quantifying location privacy* 2011 IEEE Symposium on Security and Privacy,
[141]. Simmhan, Y. L., Plale, B., & Gannon, D. (2005). A survey of data provenance in e-science. *SIGMOD Record*, *34*(3), 31-36. https://doi.org/https://doi.org/10.1145/1084805.1084812
[142]. Solove, D. J. (2006). A taxonomy of privacy. *University of Pennsylvania Law Review*, *154*(3), 477-560. https://doi.org/https://doi.org/10.2307/40041279
[143]. Standards, N. I. o., & Technology. (2019). *Recommendation for block cipher modes of operation: Methods for format-preserving encryption (NIST SP 800-38G, Update 1)*.
[144]. Standards, N. I. o., & Technology. (2020a). *Recommendation for key management—Part 1: General (NIST SP 800-57pt1r5)*.
[145]. Standards, N. I. o., & Technology. (2020b). *Zero Trust Architecture (SP 800-207)*.
[146]. Subashini, S., & Kavitha, V. (2011). A survey on security issues in service delivery models of cloud computing. *Journal of Network and Computer Applications*, *34*(1), 1-11. https://doi.org/https://doi.org/10.1016/j.jnca.2010.07.006
[147]. Subrato, S., & Md, N. (2024). The role of perceived environmental responsibility in artificial intelligence-enabled risk management and sustainable decision-making. *American Journal of Advanced Technology and Engineering Solutions*, *4*(04), 33-56. https://doi.org/10.63125/7tjw3767
[148]. Svantesson, D. J. B. (2013). Extraterritoriality and targeting in EU data privacy law: The weaker spot. *International Data Privacy Law*, *3*(4), 226-234. https://doi.org/https://doi.org/10.1093/idpl/ipt019
[149]. Sweeney, L. (2002). k-Anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*(5), 557-570. https://doi.org/https://doi.org/10.1142/S0218488502001648
[150]. Tahmina Akter, R., & Abdur Razzak, C. (2022). The Role of Artificial Intelligence in Vendor Performance Evaluation Within Digital Retail Supply Chains: A Review Of Strategic Decision-Making Models. *American Journal of Scholarly Research and Innovation*, *1*(01), 220-248. https://doi.org/10.63125/96jj3j86
[151]. Tahmina Akter, R., Debashish, G., Md Soyeb, R., & Abdullah Al, M. (2023). A Systematic Review of AI-Enhanced Decision Support Tools in Information Systems: Strategic Applications In Service-Oriented Enterprises And Enterprise Planning. *Review of Applied Science and Technology*, *2*(01), 26-52. https://doi.org/10.63125/73djw422
[152]. Tang, Y., Wang, P. P. C., Xu, Y., & Zhang, F. (2010). FADE: Secure overlay cloud storage with file assured deletion. In *Security and Privacy in Communication Networks* (pp. 380-397). https://doi.org/https://doi.org/10.1007/978-3-642-16161-2_22
[153]. Tawalbeh, L., Aldmour, I., Alasmawi, H., & Alshwaheen, A. (2022). Towards a privacy impact assessment methodology to support GDPR compliance for big data analytics. *Computer Law & Security Review*, *46*, 105728. https://doi.org/https://doi.org/10.1016/j.clsr.2022.105728
[154]. Tursunbayeva, A., Pagliari, C., Di Lauro, S., & Antonelli, G. (2021). The ethics of people analytics: Risks, opportunities, and recommendations. *Personnel Review*, *51*(3), 900-920. https://doi.org/https://doi.org/10.1108/PR-12-2019-0680
[155]. van Ooijen, I., & Vrabec, H. U. (2019). Does the GDPR enhance consumers' control over personal data? An analysis from a behavioural perspective. *Journal of Consumer Policy*, *42*(1), 91-107. https://doi.org/https://doi.org/10.1007/s10603-018-9399-7

[156]. Vassiliadis, P., Simitsis, A., & Skiadopoulos, S. (2009). A survey of extract–transform–load technology. *The VLDB Journal*, *18*(1), 125-144. https://doi.org/https://doi.org/10.1007/s00778-008-0115-y

[157]. Veale, M., & Edwards, L. (2018). Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling. *Computer Law & Security Review*, *34*(2), 398-404. https://doi.org/https://doi.org/10.1016/j.clsr.2017.12.002

[158]. Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the GDPR. *International Data Privacy Law*, *7*(2), 76-99. https://doi.org/https://doi.org/10.1093/idpl/ipx005

[159]. Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, *12*(4), 5-33. https://doi.org/https://doi.org/10.1080/07421222.1996.11518099

[160]. Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, *60*(309), 63-75. https://doi.org/https://doi.org/10.1080/01621459.1965.10480775

[161]. Weber, K., Otto, B., & Österle, H. (2009). One size does not fit all—A contingency approach to data governance. *Journal of Data and Information Quality*, *1*(1), Article 4. https://doi.org/https://doi.org/10.1145/1515693.1515696

[162]. Weitzner, D. J., Abelson, H., Berners-Lee, T., Feigenbaum, J., Hendler, J., & Sussman, G. J. (2008). Information accountability. *Communications of the ACM*, *51*(6), 82-87. https://doi.org/https://doi.org/10.1145/1409360.1409364

[163]. Yao, A. C.-C. (1982). *Protocols for secure computations* 23rd Annual Symposium on Foundations of Computer Science,

[164]. Zhang, J., Cormode, G., Procopiuc, C. M., Srivastava, D., & Xiao, X. (2017). *PrivBayes: Private data release via Bayesian networks* Proceedings of the 2017 ACM SIGMOD International Conference on Management of Data,

[165]. Zhang, Y., Deng, R. H., Xu, S., Sun, J., Li, Q., & Zheng, D. (2020). Attribute-based encryption for cloud computing access control: A survey. *ACM Computing Surveys*, *53*(4), 84. https://doi.org/https://doi.org/10.1145/3398036

[166]. Zissis, D., & Lekkas, D. (2012). Addressing cloud computing security issues. *Future Generation Computer Systems*, *28*(3), 583-592. https://doi.org/https://doi.org/10.1016/j.future.2010.12.006