International Journal of Applied Engineering & Technology

DESIGNING A PRIVACY-FIRST ARCHITECTURE FOR MULTI-ORGANIZATION DATA SHARING IN REGULATED INDUSTRIES

Ronakkumar Bathani
Sr. Data Engineer (Independent Researcher) Institute of Technology, Nirma University
ronakbathani@gmail.com

ABSTRACT

The paper discusses the ability of regulated industries to establish a privacy first design of data sharing across multiple organizations to balance legal requirements with the desire to use advanced analysis and collaboration. It is done to leave single security measures behind and describe a conceptual, reference-level architectural map that systematically implants confidentiality, integrity, and compliance at both technical and governance levels. The research employs the secondary research approach, which summarizes more recent research on confidential computing, secure multi-party analytics, cryptographic access control, blockchain-based trust fabrics, dataspaces, and collaborative intelligence in healthcare, finance, IoT, and cyber-defense domains to design common patterns and identify unresolved gaps. The paper uses critical comparison of these streams to find three fundamental results: the necessity of confidential, cross-party analytics; cross-traditional RBAC with cryptographic data-centric access control and the development of blockchain-supported governance and dataspaces as cross-organizational trust and policy planes. It is based on this that it suggests a layered privacy-first architecture, where sensitive data is kept locally, federated and confidential analytics are used, shared and auditable trust infrastructure, and a layer of policy-aware governance. These findings indicate that this type of integrated design may help mitigate compliance risk, enhance accountability, and achieve scalable and interoperable data ecosystems in highly-regulated settings, and that standard interfaces and reference implementations are necessary.

Keywords: *Data, Privacy, Control / Controls, Governance, Policy / Policies, Organization / Organizational / Organizations / Multi-organization, Cryptographic / Cryptography, Blockchain, Architecture / Architectural*

INTRODUCTION AND BACKGROUND

Controlled industries such as healthcare and finance cannot operate without exchanging data, but controlled by stringent laws, they do not have much freedom to use and transfer information. Such regulations as GDPR in Europe or HIPAA in the United States help to prevent the misuse of personal and sensitive information. Privacy-first architecture designs minimize the exposure of data and values privacy as an essential feature, not an optional feature. It involves both technical control such as encryption, fine-grained access control, and audit trails, as well as organizational regulation such as contracts, consent frameworks and clear defined roles. Such solutions as federated analytics and secure multi-party computation allow organizations to generate shared insights without transferring raw data to external institutions. By doing so, the privacy-first designs will seek to create a balance between regulatory compliance and the necessity to introduce data-driven innovation in the complex and multi-organization ecosystems.

LITERATURE REVIEW

According to Azarm-Daigle et al., (2015), interoperability gaps, lack of clear governance, and patient privacy and consent concerns are limiting cross-organizational healthcare data sharing, but it is needed to ensure continuity of care and population-level analytics. Figueiredo, (2017) redefines the concept of data sharing as an opportunity, stating that the existence of strong governance, standardization, and incentives can transform technical and ethical issues into health benefits to people and innovation potential. This is further applied to knowledge management by Asrar-ul-Haq and Anwar, (2016) who demonstrate that the perceived risk factor, trust, and organizational culture are significant factors that determine whether actors will share data and knowledge despite the existence of infrastructure.

Xia et al., (2017) introduce a blockchain-based model, BBDS, a scheme of electronic medical record sharing in cloud computing, to promote integrity, auditability, and control to patients, with distributed ledgers and cryptographic operations. Wolfert et al., (2017) show that big data in smart farming is no exception, requiring secure and interoperative data platforms that strike a balance between competitive sensitivities and shared value creation. Acquier et al., (2017) define structural paradoxes in sharing-economy platforms, in which so-called openness is accompanied by new centralization and control, heralding analogous conflicts in data platforms. According to Cheng et al., (2017) the breach of enterprise data exposes system vulnerabilities in access controls, monitoring and culture of security, and Talesh, (2018) demonstrates that cyber insurance markets influence the practices of compliance indirectly by introducing privacy and security expectations into underwriting and risk management. These works together reveal a common conflict between sharing, power, and security.

METHODOLOGY

The proposed work takes the form of secondary research to synthesize and criticize the extant technical and governance solutions to privacy-first, multi-organisation data sharing in regulated industries. The comparison of the three concepts of confidential computing, cryptographic access control, blockchain trust fabrics, and dataspaces is conducted systematically using scholarly articles, standards, and industry white papers without the limitation of one organizational case. Secondary data assists in providing extensive coverage of cross-sector practices, which would cover healthcare, finance, IoT, and cyber-defense settings and which would be unrealistic to discover empirically in a single project. It also allows the triangulation of the results between independent implementations and the discovery of convergent architectural patterns and continuing integration gaps that a single individual, context-specific study could fail to identify (Cheong *et al*. 2023). Besides, secondary analysis allows a critical interaction with officially defined protocols, reference models, and interpretations of regulatory standards, which is crucial to the creation of an abstract, generalizable privacy-first reference architecture and not a custom-built system design.

RESULTS

Need for Confidential Multi-Party Analytics

Secrecy multi-party analytics solves a structural conflict of generating insights through collaboration and the firmness of their non-disclosure obligations within regulated ecosystems. Companies have to collaboratively identify fraud, streamline supply chains, or match up cyber threats but cannot publish unprocessed data sets with personal, proprietary, or classified data.

Table 1: Techniques and Properties

Technique	Trust Assumption	Privacy Strength	Performance Overhead	Typical Use Case
Secure Multi-Party Computation	Honest-but-curious parties	High	High	Joint analytics without data pooling
Homomorphic Encryption	Honest cloud, correct implementation	Very high	Very high	Encrypted model training or scoring
Trusted Execution Environments	Trusted hardware vendor and attestation	High	Medium	Confidential analytics on public cloud
Differential Privacy	Trusted curator for noise calibration	Medium-high	Low-medium	Release of aggregate statistics and models

Confidential analytics is thus based on cryptographic and hardware-based designs that mathematically limit information leakage in computation and storage (Urciuoli and Hintsa, 2017).

Secure multi-party computation is a protocol that allows multiple parties to collectively compute functions on distributed inputs without necessarily knowing those inputs individually. Homomorphic encryption supports restricted computation on ciphertexts, thus no processing in plaintext form is ever seen. Trusted execution environments provide hardware-isolated enclaves in which code and data are isolated even to special purpose system administrators (Shinde *et al*. 2017). Practically, these primitives are coordinated by data clean rooms, federated query engine, and privacy-preserving model training pipeline.

Table 2: Risk–Control Mapping

Risk Type	Example Scenario	Mitigating Control	Residual Risk Level
Re-identification	Linking outputs to individuals	Differential privacy	Medium
Insider data snooping	Analyst abusing access	MPC / TEEs	Low-medium
Cloud operator exposure	Provider reading data in clear	Homomorphic encryption / TEEs	Low
Cross-party inference	Partner infers competitor’s strategy	Output access policies + DP	Medium

The query planners break down analytics into local tasks which only exchange encrypted aggregates or masked intermediate results. Calibrated noise is then introduced to outputs as part of differential privacy, making them inattackable by re-identification or differencing attacks. Formal guarantees are used in controlled industries where competition cannot be able to recreate sensitive properties using model parameters or logs (Gjerdrum *et al*. 2023). Purpose limitation, retention limits, and consent scopes are imposed by policy-aware runtimes on each analytic invocation. In this way, confidential multi-party analytics does not become a pleasant addition that can be optional, but a set of architectural principles that enable legal, large-scale sharing of data under current regulatory frameworks.

Blockchain as a Cross-Organizational Trust Fabric

Blockchain is a common, tamper evident trust substrate, coordinating permissions and accountability among independent organizations. Multi-party data sharing does not permit the capacity of one institution to be a universally trusted policy administrator and log keeper.

Table 3: Ledger Design Choices

Dimension	Option A: Permissioned	Option B: Permissionless	Recommended for Regulated Use
Validator Identity	Known organizations	Pseudonymous nodes	Permissioned
Consensus	BFT / RAFT	PoW / PoS	BFT / RAFT
Throughput	High	Variable	High
Governance	Consortium agreement	Protocol community	Consortium

Permissioned blockchains solve this by distributing ledgers with a set of vetted participants, and reach Byzantine fault-tolerant state transition consensus (Zhang *et al*. 2017). Smart contracts contain access control policy, consent requirements, and data usage requirements as code and can be executed. Every transgression made by cross-organizational access request is an appraisal of these contracts prior to data circulation. The decisions and related cryptographic evidence are permanently stored and auditable records of either conformity or nonconformity are formed. The identity layers attach organizational identity, device certificates and role credentials to on-chain representations.

Table 4: On-Chain vs Off-Chain Split

Function	On-Chain Component	Off-Chain Component	Rationale
Policy registration	Smart contracts	Policy authoring tools	Immutability and audit
Consent logging	Transaction records	Consent UI / identity provider	Verifiable evidence

Data storage	Hash and locator	Encrypted object store or dataspace node	Scalability and privacy
Key management events	Key ID and status	HSM / KMS handling actual keys	Confidentiality of secrets

Such mapping facilitates attribute-based/role-based authorization, which coordinates the release of decryption keys and delegation of capabilities across domains. In case of IoT and cyber-physical systems, blockchain verifies parties and message hashes, which ensure end-to-end provenance and integrity. Enforcement of the policies can be audited by regulators and auditors, without relying on internal logs that are not verifiable or unilateral reporting. Together with off-chain storage and encryption, the ledger regulates who has access to what views or aggregates on what reasons (Wolf and Serpanos, 2017). Therefore, blockchain transitions to a programmable plane of governance which operationalizes zero trust assumptions in multi-organization, privacy-critical data architectures.

Cryptographic Access Control beyond Traditional RBAC

Role-based access control is traditional and attaches access permissions to fixed organizational roles, which is not very dynamic in cross-organization data sharing where context, attributes, and fine-grained obligations are very important. In cryptographic access control, server-side checks are substituted or supplemented with cryptographically enforceable policies that bind decryption power to user or service attributes, roles or context constraints (Premarathne *et al*. 2016). Attributes Ciphertexts in attribute-based encryption and role-based encryption schemes have policy logic embedded in them, with only those principals possessing matching cryptographic keys being able to decrypt plaintext.

Table 5: Model Comparison

Model	**Policy Location**	**Enforcement Point**	**Revocation Complexity**	**Cross-Org Scalability**
Classic RBAC	Application server	Application layer	Low	Medium
Attribute-Based Access	Policy engine	Token / gateway	Medium	High
Role-Based Encryption	Ciphertext metadata	Decryption operation	Medium-high	High
Attribute-Based Encryption	Ciphertext metadata	Decryption operation	High	Very high

This moves the trust boundary out of application servers to verifiably key distributions that are mathematically provable, and minimizing insider abuse and broken policy engines. The key management authorities may be delegated or hierarchically structured in multi-tenant or federated cloud environments, allowing the delegated administration without putting all control in one place. Separation-of-duty, least-privilege, temporal validity, and purpose restriction may be coded in fine-grained policies, applied at decryption time, and not just requested time (Yang *et al*. 2016). Attached hardware security modules or trusted execution environments will additionally safeguard key material and policy interpreters against system-level attackers.

Table 6: Data Object Lifecycle Controls

Lifecycle Stage	**Control Objective**	**Cryptographic Mechanism**	**Governance Artifact**
Creation	Bind data to policy	ABE/RBE encryption at write	Data classification tag
Storage	Prevent unauthorized at-rest use	Key wrapping, HSM-backed keys	Key management policy
Sharing	Enforce least privilege	Re-encryption, scoped key grants	Data sharing agreement
Deletion	Ensure effective cryptographic erase	Key revocation / destruction	Retention schedule

Cryptographic access control therefore offers end-to-end, data-centric protection that is portably relied upon, and support privacy-preserving sharing in a more robust control than ordinary role-based controls.

Dataspaces and Collaborative Intelligence as Governance Layers

Dataspaces represent a conceptualization of sharing data as a federated ecosystem in which autonomous participants share semantically described resources, subject to the rules of shared governance, as opposed to giving them up to a single centralized repository. A dataspace infrastructure offers metadata catalogs, semantic interoperability services and policy-aware access brokers to mediate discovery and constrained use of heterogeneous assets (Garofalaki *et al*. 2017). The governance is carried out as a stacked control plane that articulates data usage policy, consent constraints, jurisdictional boundaries and contractual requirements as machine readable rules. Trusted integrated knowledge dataspaces scale this paradigm to sensitive spaces, incorporating anonymization, limitations of purpose and provenance into query and integration pipelines.

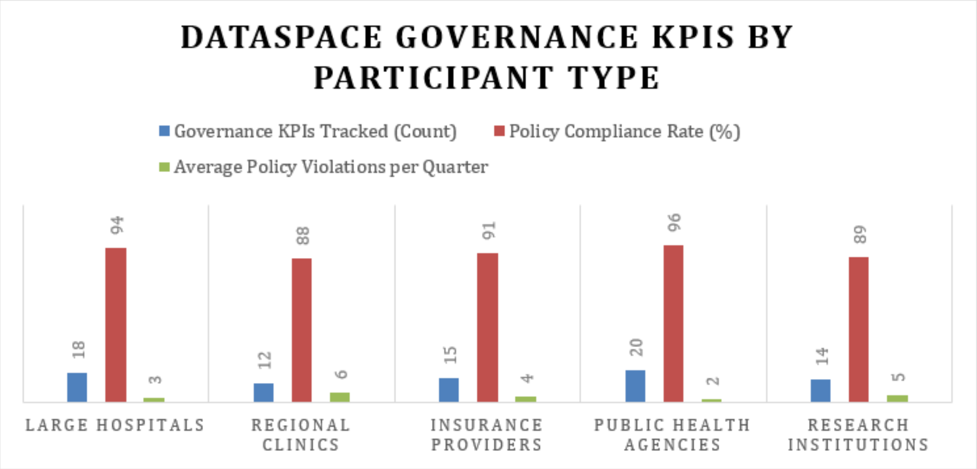

Figure 1: Dataspace Governance KPIs by Participant Type

Organized distributed analytics, model-training, and inference across multiple nodes on the foundation of these foundations build on collaborative intelligence, and keep data local and controlled. Policy engines compare the requests to the ownership, to consent and regulatory context and dynamically route the computations or generate derived views (Buczak and Guven, 2015). Lineage and provenance records capture the history of combining datasets, features, and models, which allow establishing accountability, reproducibility, and impacts of downstream decisions.

According to this perspective, dataspaces and collaborative intelligence constitute governance layers mediating between semantics, policy and accountability, above raw transport and storage, to provide scalable, compliant, multi-organization data ecosystems.

Absence of an Integrated Privacy-First Reference Architecture

Current contributions are more likely to focus on the individual aspects of the privacy-preserving data sharing, e.g. confidential computing, blockchain-based access control, or trusted dataspaces, but seldom combine them into a comprehensive reference architecture of regulated multi-organization contexts.

Table 7: Layered Stack Overview

Layer	Main Responsibility	Key Technologies
Data protection	Confidentiality and integrity	Encryption, TEEs, MPC, DP
Trust and audit	Shared state and logging	Permissioned ledger, signed events
Governance / policy	Express and evaluate rules	Policy engine, consent manager
Application / analytics	Domain logic and models	Dashboards, AI pipelines, APIs

Cryptographic primitives, confidential enclaves and secure multi-party computation have robust data-in-use guarantees, but are commonly not standardized to governance, consent and provenance layers across industries (Sinha *et al*. 2015). On the other hand, governance-based models and dataspaces establish the policies, roles and responsibility, but presuppose standard security measures and do not directly include end-to-end cryptographic implementation. Blockchain fabrics and zero trust models provide auditable, decentralized policy enforcement, but the combination with privacy enhancing technology and industry-specific compliance processes is ad hoc and pattern-based. No adopted architectural blueprint has seen the systematic arrangement of identity, key management, confidential execution, consent management, and policy reasoning and auditing into a coherent, privacy-first stack.

Table 8: Requirement–Capability Matrix

Requirement	Supporting Capability	Primary Layer
Regulatory compliance	Policy engine, consent, retention controls	Governance / policy
Strong confidentiality	Encryption, MPC, TEEs, DP	Data protection
Cross-org accountability	Immutable logs, provenance, audits	Trust and audit
Interoperability and scalability	Dataspace connectors, standard schemas	Governance / application

It would be characterized by such an architecture, which identifies control planes, trust anchors, references interfaces, and lifecycle processes within ingestion, processing, sharing, and deletion across organizations (Gürcan and Berigel, 2018). Lack of such an integrated perspective disintegrates implementations, adds compliance risk, and makes reusable, certified solutions difficult to achieve, and the need to use a single privacy-first reference architecture to regulated multi-organization data sharing is explicitly stated.

Suggested Design

The proposed privacy-first framework must consist of data-centric protection, decentralized trust, and robust governance across organizations in controlled industries. Fundamentally, sensitive information is stored in the environment of every organization and is encrypted, encrypted with role or attribute-based encryption, and encrypted in-use with the help of confidential computing to protect sensitive data, and with secure multi-party computation or federated learning to prevent the movement of raw data across organizations. Shared policy enforcement and immutable audit trails on all access to data, consent decision, and key operations can be supported by a consortium trust layer, such as through permissioned ledgers, and based on verifiable identities, to answer zero trust assumptions across parties. Most importantly, above these, a dataspace-inspired governance layer standardizes the semantics, presents policy-sensitive APIs and integrates consent, purpose limitation and retention rules into query planning and model deployment, such that compliance is built in, not required by hand.

- Local-first storage with encrypted, policy-bound data objects.
- Federated or confidential analytics instead of central data pooling.
- Shared trust and audit plane for cross-organization access events.
- Dataspace-based governance for semantics, consent, and lifecycle control.

DISCUSSION

The results show a landscape that is rich in technology and disjointed making adaptation of the same in regulated sectors not easy in the real world. Multi-party analytics that use confidentiality, cryptographic access control, and blockchain-based trust fabrics each mitigate individual risks but add complexity to key management, interoperability, and performance that many organizations will have difficulty operationalizing (Gjerdrum *et al*. 2017). Collaborative intelligence and dataspaces have attractive governance abstractions, although they usually presuppose mature metadata practices and uniform policy vocabularies, which seldom exist in heterogeneous consortia. The proposed privacy-first architecture partially addresses this fragmentation with layered confidential calculation, decentralized trust and governance, but it runs the risk of becoming overly idealized unless these integration costs, legacy system and sector-specific constraints are explicitly addressed. Lack of interoperability between cryptographic enforcement, ledger-based policy planes and dataspace governance can keep the architecture just a conceptual blueprint not a deployable pattern, highlighting the importance of reference implementations, performance benchmarks and regulatory validation across a variety of industries.

CONCLUSION

This paper has revealed that existing strategies of sharing data between multi-organizations in regulated industries are advanced but disconnected and lack essential integration and governance considerations. Each of the confidential multi-party, cryptographic access, blockchain-based trust fabrics, and dataspaces address particular privacy and compliance threat, and are rarely operated as a coherent architecture. Through combining these streams, the paper suggests a privacy-first reference architecture where sensitive data are retained locally, trust is externalized through infrastructures which are sharing and auditable, and policy and consent enforcement are implemented within all analytic interactions, instead of at system boundaries. This design re-imagines privacy as a compliance (rather than compliance) after-thought into the key organizing concept of technical, organizational, and regulatory alignment. The work thus provides a critical perspective of the current solutions, as well as a blueprint of structured plans to construct compliant, interoperable and future ready data ecosystems in regulated sectors.

REFERENCES

Acquier, A., Daudigeos, T. and Pinkse, J., 2017. Promises and paradoxes of the sharing economy: An organizing framework. *Technological Forecasting and Social Change*, *125*, pp.1-10.

Asrar-ul-Haq, M. and Anwar, S., 2016. A systematic review of knowledge management and knowledge sharing: Trends, issues, and challenges. *Cogent business & management*, *3*(1), p.1127744.

Azarm-Daigle, M., Kuziemsky, C. and Peyton, L., 2015. A review of cross organizational healthcare data sharing. *Procedia Computer Science*, *63*, pp.425-432.

Buczak, A.L. and Guven, E., 2015. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications surveys & tutorials*, *18*(2), pp.1153-1176.

Cheng, L., Liu, F. and Yao, D., 2017. Enterprise data breach: causes, challenges, prevention, and future directions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *7*(5), p.e1211.

Cheong, H.I., Lyons, A., Houghton, R. and Majumdar, A., 2023. Secondary qualitative research methodology using online data within the context of social sciences. *International Journal of Qualitative Methods*, *22*, p.16094069231180160.

Figueiredo, A.S., 2017. Data sharing: convert challenges into opportunities. *Frontiers in public health*, *5*, p.327.

Garofalaki, Z., Kallergis, D., Katsikogiannis, G. and Douligeris, C., 2017. A Policy-Aware Model for Intelligent Transportation Systems. *arXiv preprint arXiv:1706.04803*.

Gjerdrum, A.T., Pettersen, R., Johansen, H.D. and Johansen, D., 2017, April. Performance of Trusted Computing in Cloud Infrastructures with Intel SGX. In *CLOSER* (pp. 668-675).

Gürcan, F. and Berigel, M., 2018, October. Real-time processing of big data streams: Lifecycle, tools, tasks, and challenges. In *2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)* (pp. 1-6). IEEE.

Premarathne, U., Abuadbba, A., Alabdulatif, A., Khalil, I., Tari, Z., Zomaya, A. and Buyya, R., 2016. Hybrid cryptographic access control for cloud-based EHR systems. *IEEE Cloud Computing*, *3*(4), pp.58-64.

Shinde, S., Le Tien, D., Tople, S. and Saxena, P., 2017, February. Panoply: Low-TCB Linux Applications With SGX Enclaves. In *NDSS*.

Sinha, R., Rajamani, S., Seshia, S. and Vaswani, K., 2015, October. Moat: Verifying confidentiality of enclave programs. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (pp. 1169-1184).

Talesh, S.A., 2018. Data breach, privacy, and cyber insurance: How insurance companies act as “compliance managers” for businesses. *Law & Social Inquiry*, *43*(2), pp.417-440.

Urciuoli, L. and Hintsa, J., 2017. Adapting supply chain management strategies to security–an analysis of existing gaps and recommendations for improvement. *International Journal of Logistics Research and Applications*, *20*(3), pp.276-295.

Wang, H., Sua, L.S. and Alidaee, B., 2024. Enhancing supply chain security with automated machine learning. *arXiv preprint arXiv:2406.13166*.

Wolf, M. and Serpanos, D., 2017. Safety and security in cyber-physical systems and internet-of-things systems. *Proceedings of the IEEE*, *106*(1), pp.9-20.

Wolfert, S., Ge, L., Verdouw, C. and Bogaardt, M.J., 2017. Big data in smart farming–a review. *Agricultural systems*, *153*, pp.69-80.

Xia, Q., Sifah, E.B., Smahi, A., Amofa, S. and Zhang, X., 2017. BBDS: Blockchain-based data sharing for electronic medical records in cloud environments. *Information*, *8*(2), p.44.

Yang, K., Han, Q., Li, H., Zheng, K., Su, Z. and Shen, X., 2016. An efficient and fine-grained big data access control scheme with privacy-preserving policy. *IEEE Internet of Things Journal*, *4*(2), pp.563-571.

Zhang, J., Liu, Q., Hu, Z., Lin, J. and Yu, F., 2017. A multi-server information-sharing environment for cross-party collaboration on a private cloud. *Automation in Construction*, *81*, pp.180-195.