A Blockchain-Based Secure Data Sharing Framework for Distributed Computing Environments

Breoks Cistra
Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA.
brookscastro97@missouri.edu

Abstract

The sharing of data across distributed computing environments remains a persistent challenge because participants often operate under heterogeneous administrative domains, inconsistent security policies, and different levels of technical capability. Centralized data sharing platforms provide functional convenience but introduce single points of failure, governance bottlenecks, and concerns about unauthorized access or manipulation. Blockchain technology offers an alternative by providing a tamper-evident distributed ledger that can coordinate identity, access policy, provenance, and audit across multiple parties without requiring a single trusted intermediary. This paper presents a system-level framework for blockchain-based secure data sharing in distributed computing environments. The framework is organized around a layered architecture that separates data storage, access control, consensus, contract execution, and governance. The discussion emphasizes structural trade-offs rather than low-level cryptographic operations. It examines how permissioned and permissionless configurations affect confidentiality, scalability, robustness, and accountability. The paper also addresses incentive design, deployment sustainability, fairness, regulatory alignment, and long-term operational considerations. Through cross-domain analysis of healthcare, Internet of Things, energy trading, and scientific collaboration contexts, the paper identifies architectural principles that can guide the construction of resilient and socially legitimate data sharing infrastructures. It concludes that blockchain-based data sharing is not a universal substitute for institutional trust, but rather a coordination mechanism whose effectiveness depends on careful architectural configuration, inclusive governance, and policy coherence.

Keywords

blockchain, secure data sharing, distributed computing, smart contracts, decentralized governance, data sovereignty, interoperability.

1. Introduction

Distributed computing environments now span cloud data centers, edge nodes, institutional clusters, mobile devices, and Internet of Things deployments. These environments are characterized by administrative fragmentation, dynamic resource availability, and heterogeneous trust assumptions. When organizations and individuals share data across such environments, they must simultaneously protect confidentiality, preserve data integrity, establish provenance, and demonstrate accountability. Traditional access control mechanisms usually depend on centralized identity providers and policy servers. While these mechanisms simplify administration, they create concentrated points of failure and require all parties to accept the authority of a platform operator. Blockchain technology introduces an alternative coordination model in which a replicated ledger provides a shared, append-only record of events without a central mediator. The original formulation of Bitcoin demonstrated that

mutually distrusting participants could maintain consensus over transaction ordering [1]. Subsequent systems extended this concept with programmable smart contracts that execute conditional logic across a decentralized network [2]. These capabilities have stimulated interest in blockchain as a foundation for secure data sharing, because data access decisions, consent records, and audit trails can be expressed as on-chain state transitions.

A system-level perspective is essential when applying blockchain to data sharing in distributed computing. The technology cannot be treated merely as a cryptographic database; it is a socio-technical infrastructure that interacts with organizational policy, legal frameworks, economic incentives, and operational constraints. Privacy-sensitive data sharing requires sophisticated approaches such as decentralized access control and user-mediated consent [3]. In Internet of Things settings, blockchain-connected gateways and sensors face severe resource constraints, demanding lightweight protocols and careful placement of trust boundaries [4]. Optimized blockchain designs have been proposed specifically for constrained device networks, balancing verification overhead with security requirements [5]. Practical Byzantine fault tolerance provides a foundation for consortium settings where a known set of participants can tolerate a bounded number of Byzantine failures [6]. These foundational works reveal that blockchain-based data sharing is not a single monolithic design, but a family of architectural choices shaped by the number of participants, the sensitivity of data, the available infrastructure, and the required performance.

2. Foundational Challenges and System Requirements

Distributed data sharing presents challenges that extend beyond basic encryption. Participants may not know each other directly, may operate under different legal regimes, and may have conflicting interests in how data are used. A centralized data exchange can enforce a uniform policy, but it becomes an attractive target for attack and a potential source of unfair market power. Blockchain systems address part of this problem by providing a replicated state machine whose transitions can be jointly validated. Enterprise-oriented modular blockchain architectures demonstrate how membership services, consensus mechanisms, and smart contract execution can be decoupled to support multi-party applications [7]. This modularity is particularly valuable when different participants require different levels of access and different auditing obligations. In many large-scale data sharing scenarios, storing raw datasets directly on-chain is neither efficient nor privacy-preserving. Content-addressed peer-to-peer storage systems have therefore been combined with blockchain ledgers, allowing the ledger to hold integrity references while raw data remain in distributed storage [8]. This separation reduces storage pressure and supports selective disclosure, but it also creates a dependency between ledger availability and external storage availability.

The system requirements for secure data sharing include authenticity, non-repudiation, access control, auditability, and confidentiality. Authenticity can be strengthened by cryptographic certifications that bind data objects to issuer identity and metadata [9]. In healthcare, blockchain-based platforms have been proposed to let patients manage permissions for electronic health records while providing an immutable log of access events [10]. Cross-domain comparative analyses have produced taxonomies of blockchain-based systems that clarify design dimensions such as consensus openness, smart contract capability, and data storage placement [11]. These dimensions matter because a framework that works for medical records may be unsuitable for high-frequency sensor data or financial transactions. A general framework must therefore define interfaces and separation points rather than prescribing a single data structure or consensus algorithm. It must also support data minimization, consent

revocation, role separation, and the ability to prove compliance to external auditors without exposing unnecessary information. These requirements cannot be satisfied by cryptography alone; they demand a carefully designed interaction among identity systems, contract logic, data stores, and governance mechanisms.

3. Architectural Principles of the Framework

The proposed framework is organized into five interconnected layers: a data plane, a control plane, a ledger plane, a contract plane, and a governance plane. The data plane contains distributed storage, edge caches, and domain-specific databases. The control plane manages identity, authentication, and policy enforcement points. The ledger plane maintains the replicated record of access events, policy commitments, and data provenance. The contract plane executes logic that governs data sharing, such as evaluating access requests, checking consent, and triggering notifications. The governance plane includes human decision-making processes, dispute resolution, and policy update procedures. This layered separation prevents changes in one plane from destabilizing the others. For example, a storage migration in the data plane should not require a consensus protocol upgrade, and a smart contract update should not rewrite historical access logs. Such modularity has parallels in enterprise blockchain architectures where the ordering of transactions is separated from contract execution [7]. Smart contracts on platforms such as Ethereum illustrate the expressive power of on-chain logic, but they also show therisks of treating executable code as an unconditional institutional commitment. Contract logic must be designed for upgradeability, dispute pauses, and human override where legal obligations conflict with automated execution. In the proposed framework, the governance plane is responsible for defining the conditions under which contract behavior can be paused or amended. This capability is not a weakness; it is a recognition that distributed data sharing operates within legal orders that may require the right to be forgotten, mandatory breach disclosure, or lawful interception. A purely immutable contract would conflict with such obligations. The architecture therefore treats immutability as a property of historical ledger entries and audit records, not as an absolute constraint on future operational rules.

The layered design also clarifies where trust boundaries are redistributed rather than removed. A permissioned consortium can limit validator participation to institutions that have completed know-your-customer and contractual onboarding. This configuration improves throughput, reduces energy consumption, and aligns with regulatory expectations, but it also concentrates validation power. Permissionless networks provide stronger censorship resistance and broader participation, but they impose replication costs on every node and make confidential data leakage harder to control. The trade-off between these configurations has been characterized as a choice between resilience through open participation and accountability through identified participants [12]. Many real deployments adopt intermediate configurations, such as permissioned ledgers with publicly verifiable anchors or permissionless networks with privacy-preserving client channels. The proposed framework accommodates these variations by treating consensus participation as a governance parameter rather than an architectural constant.

Identity management is a critical control plane function. Early blockchain systems used public keys as pseudonymous identities, which supports integrity but not regulated data sharing. To meet legal requirements, the framework supports decentralized identifiers and verifiable credentials that allow participants to prove attributes without revealing unnecessary personal data [13]. A data provider can issue a machine-readable credential stating that a requester

belongs to an accredited research institution or has completed ethics training. The smart contract can evaluate the credential and record only its validity status on-chain. This approach preserves auditability while limiting the exposure of personal identity data. It also enables selective disclosure, where a requester proves a bounded set of attributes instead of transferring a complete identity profile. The verifiable credentials are stored in the control plane, while the ledger stores revocation state and credential schemas.

Off-chain computation and storage raise the question of how to trust that data processing follows the agreed contract. Cryptographic commitments alone cannot verify that a remote node executed a confidential computation correctly. Trusted execution environments provide one enforcement mechanism by binding computation results to hardware attestations that can be recorded on-chain [14]. Although trusted execution environments have been subject to side-channel attacks and hardware vendor dependencies, they offer a practical middle ground for large datasets that cannot be processed on-chain. The framework treats trusted execution environments as an optional security substrate rather than a core requirement, because different domains have different threat models. In highly regulated healthcare settings, hardware enclaves may be acceptable only when combined with contractual guarantees and independent auditing. In lower-sensitivity scientific collaborations, simple cryptographic checksums and access logging may suffice.

4. Security and Privacy Trade-offs

Secure data sharing must confront a central tension: auditability often requires recording metadata about who accessed what data, while privacy requires minimizing such metadata. Blockchain ledgers make records permanent and publicly verifiable, which can amplify this tension if access events are stored without protection. The framework therefore distinguishes between data objects, access requests, access grants, and audit events. Data objects remain off-chain, while access grants and audit events are recorded on-chain only in protected form. When stronger anonymity is required, zero-knowledge proofs can be used to demonstrate that a transaction satisfies a policy without revealing the underlying inputs [15]. This technique allows a system to prove that an access request is authorized without disclosing the requester identity or the specific data attributes. However, zero-knowledge proof generation remains computationally expensive, and its practical use may be limited to infrequent, high-value access decisions.

In Internet of Things environments, the resource constraints of sensors and gateways introduce additional privacy risks. A smart home system using blockchain has been proposed to manage device commands and access policies locally while using the ledger for audit and anomaly detection [16]. Such designs illustrate that blockchain need not be embedded in every sensor. Instead, a gateway can aggregate device transactions and submit periodic commitments, reducing overhead while preserving a verifiable history. The trade-off is that the gateway becomes a more attractive attack target. The framework therefore requires that gateways be hardened, monitored, and subject to tamper-evident logging. It also allows for the use of federated learning coordinated by blockchain, in which local model updates are shared instead of raw data [17]. This approach reduces privacy leakage while still enabling global model improvement, but it introduces concerns about gradient inversion and model poisoning. Consequently, the governance plane must define acceptable participation thresholds and anomaly response procedures.

Healthcare illustrates the importance of privacy-preserving record sharing. The MedRec prototype demonstrated that blockchain can coordinate access permissions for electronic

health records without storing clinical data on-chain [10]. Subsequent work built on this idea to link blockchain components with existing clinical data standards such as FHIR [18]. The framework generalizes this pattern by requiring interoperability adapters between domain-specific data schemas and the blockchain-backed control plane. Such adapters translate clinical, energy, or scientific metadata into a common access policy language. They do not force all domains to adopt the same data model. Instead, they create a controlled vocabulary for expressing sharing conditions, such as purpose limitation, retention duration, and re-identification constraints. This avoids the common failure mode in which a blockchain platform imposes an overly generic data model that cannot capture domain-specific regulatory obligations.

5. Cross-Domain Case Illustrations and Structural Comparisons

A comparison of healthcare, Internet of Things, energy trading, and scientific collaboration reveals recurring architectural patterns. In healthcare, data sharing is driven by patient consent, institutional responsibility, and strict regulatory mandates. The blockchain ledger serves primarily as an audit and consent coordination mechanism. In an Internet of Things deployment, the ledger coordinates device identity, access tokens, and periodic state commitments, while most data remain in local gateways or edge storage. An architecture for scalable access management in Internet of Things has been proposed that uses blockchain to maintain an access control list and reduce reliance on a central broker [19]. This design improves resilience against device spoofing and central server failure, but it must handle network partitions and delayed transactions. The proposed framework accounts for these conditions by requiring asynchronous conflict resolution and idempotent access operations.

In energy trading, blockchain can support local electricity markets in which prosumers trade energy autonomously [20]. Unlike healthcare, energy trading involves frequent low-value transactions and time-sensitive settlement. The framework responds by placing market clearing and high-frequency price discovery off-chain, while the blockchain records final settlement, net metering commitments, and dispute evidence. This separation prevents the ledger from becoming a throughput bottleneck and allows existing grid operators to remain responsible for physical delivery. Fairness becomes a central concern because automated markets can disadvantage participants with limited forecasting ability or intermittent renewable generation. The governance plane can impose default safety parameters, price corridors, and mandatory fallback contracts to prevent exploitative behavior. The structural lesson is that blockchain should not replace existing energy market institutions wholesale but should provide a transparent dispute and settlement layer that enhances their accountability.

Scientific collaboration provides a different set of constraints. Research datasets may be large, versioned, and licensed under heterogeneous conditions. Blockchain can record dataset hashes, curation events, and citation provenance without transferring the datasets themselves. Auditable storage designs have been proposed for Internet of Things data that can also support scientific workflows [21]. In multi-institution research consortia, a permissioned ledger can manage access credentials, publication embargoes, and compliance with funder mandates. The main challenge is not transaction throughput but long-term preservation and metadata evolution. The framework therefore treats dataset identifiers, schema definitions, and policy templates as versioned ledger objects that can be updated through governance-approved processes. This versioning ensures that historical provenance remains interpretable even when data formats change.

These cases show that blockchain-based data sharing is not a single product but an architectural pattern whose value depends on the alignment between ledger guarantees and domain-specific institutional needs. Healthcare emphasizes privacy and regulatory auditability, Internet of Things emphasizes lightweight resilience, energy trading emphasizes fair settlement, and scientific collaboration emphasizes provenance and preservation. The framework accommodates these differences by making the ledger plane generic and pushing domain logic into contract and governance layers. This separation prevents feature creep in the core consensus mechanism and allows each domain to evolve its own policy logic without undermining cross-domain interoperability.

6. Governance, Incentives, and Deployment Sustainability

A blockchain-based data sharing system cannot remain operational without a sustainable governance model. Open-source infrastructure often suffers from free-rider problems, where many participants benefit from a shared ledger but few contribute resources. In permissioned consortia, membership fees and contractually defined operating responsibilities can mitigate this problem. In permissionless systems, economic incentives such as transaction fees and block rewards support validators, but they can also concentrate wealth and influence. The governance plane must therefore define mechanisms for validator rotation, fee adjustment, and emergency response. It must also establish a clear legal identity for the consortium or foundation that owns the codebase and manages intellectual property. Without such institutional anchoring, deployments risk becoming orphaned when key contributors depart or when participating organizations change priorities.

Regulatory alignment is another sustainability condition. Blockchain systems are often described as a form of regulatory technology that can transform legal compliance from periodic human audits to continuous automated verification [22]. This transformation is attractive but incomplete. Smart contracts can enforce rules that are unambiguous and machine-readable, but many legal obligations involve contextual judgment, proportionality, and exceptions. The framework therefore positions blockchain as an evidence and enforcement layer rather than a substitute for legal reasoning. Contracts may record that an access request was approved according to a coded policy, but human adjudication remains necessary for disputes involving ambiguous consent or conflicting rights. The governance plane must include a multi-stakeholder dispute resolution process that can interpret ledger evidence within the applicable legal framework.

Fairness in data sharing also requires attention to access asymmetries. Large institutions with substantial computational resources may be able to operate validator nodes more easily than small research groups or individual data subjects. If access decisions are based solely on token holdings or computational stake, the system may reproduce existing inequalities. The framework can mitigate this by separating validity rights from voting rights, allowing passive data contributors to delegate governance authority to trusted representatives. It can also adopt identity-weighted voting where verified institutional membership is balanced against economic stake. These mechanisms are contentious and require continuous adjustment, but they are essential for maintaining legitimacy in public-interest data ecosystems. A purely technical system that ignores power asymmetries may achieve cryptographic security while failing to achieve social acceptability.

Deployment sustainability also involves cost management. Full replication of all data on every node is expensive and often unnecessary. The framework reduces storage pressure by storing only hashes, metadata, and access events on-chain, while raw data reside in distributed

storage. This separation introduces an availability dependency on external storage providers, which can be managed through replication agreements, erasure coding, and periodic integrity audits. The ledger can store multiple redundant content hashes and monitor availability challenges. In this way, the blockchain functions as a coordination layer for storage guarantees without becoming the storage system itself. Such hybrid designs have been explored in content-addressed file systems and have shown that the combination of off-chain storage with on-chain integrity references can scale to large scientific datasets [8].

Finally, operational monitoring and incident response are critical. A blockchain system can record events immutably, but it cannot prevent all harmful actions before they occur. Malicious or erroneous smart contract operations may cause unauthorized data exposure or denial of service. The framework therefore requires a security operations capability that monitors ledger state, contract invocations, and gateway behavior for anomalies. When an incident occurs, the governance plane must be able to suspend affected contracts, revoke credentials, and initiate forensic analysis. This capability must itself be subject to checks and balances so that emergency powers cannot be abused by a single operator. A multi-signature emergency council, audited pause mechanisms, and periodic public incident reporting can help maintain both security and accountability.

7. Conclusion

This paper has presented a system-level framework for blockchain-based secure data sharing in distributed computing environments. The framework separates data storage, access control, consensus, contract execution, and governance into distinct planes. This separation allows domain-specific requirements to be addressed without overloading the core ledger. It also clarifies the distribution of trust, the handling of confidentiality, and the institutional mechanisms needed for long-term sustainability. The analysis has shown that blockchain is not a universal answer to all data sharing problems. It is most effective when used as a coordination and audit layer that integrates with existing institutional processes rather than replacing them.

The cross-domain examination of healthcare, Internet of Things, energy trading, and scientific collaboration demonstrated that different sectors require different balances among privacy, throughput, fairness, and provenance. A permissioned healthcare consortium may prioritize regulatory alignment and patient consent, while an Internet of Things network may prioritize lightweight gateway aggregation and resilience. Energy markets emphasize fair settlement and dispute transparency, whereas scientific collaboration values long-term provenance and versioned metadata. The proposed framework accommodates these differences by treating consensus membership, identity verification, privacy-preserving computation, and governance procedures as configurable parameters.

The paper has also emphasized that security cannot be separated from governance and policy. A technically sound ledger can fail socially if it concentrates authority, ignores regulatory ambiguity, or proves economically unsustainable. Fairness requires explicit mechanisms for participation, representation, and dispute resolution. Deployment requires careful attention to storage dependencies, operational monitoring, and incident response. Blockchain-based data sharing will continue to evolve as trusted execution environments, zero-knowledge proofs, and decentralized identifiers mature. However, these technical advances will only realize their potential when embedded in institutional arrangements that respect both the power and the limits of distributed systems. The framework offered here is intended to guide researchers,

system architects, and policymakers in constructing data sharing infrastructures that are secure, resilient, and aligned with broader public values.

References

1. Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. White paper.

2. Wood, G. (2014). Ethereum: A secure decentralised generalised transaction ledger. Ethereum Project Yellow Paper, 151, 1-32.

3. Zyskind, G., Nathan, O., & Pentland, A. (2015). Decentralizing privacy: Using blockchain to protect personal data. 2015 IEEE Security and Privacy Workshops, 180-184.

4. Christidis, K., & Devetsikiotis, M. (2016). Blockchains and smart contracts for the internet of things. IEEE Access, 4, 2292-2303.

5. Huh, S., Cho, S., & Kim, S. (2017). Managing IoT devices using blockchain platform. 2017 19th International Conference on Advanced Communication Technology (ICACT), 464-467.

6. Castro, M., & Liskov, B. (1999). Practical Byzantine fault tolerance. Proceedings of the Third Symposium on Operating Systems Design and Implementation, 173-186.

7. Androulaki, E., Barger, A., Bortnikov, V., Cachin, C., Christidis, K., De Caro, A., Enyeart, D., Ferris, C., Laventman, G., Manevich, Y., Muralidharan, S., Murthy, C., Nguyen, B., Sethi, M., Singh, G., Smith, K., Sorniotti, A., Stathakopoulou, C., Vukolić, M., ... Yellick, J. (2018). Hyperledger Fabric: A distributed operating system for permissioned blockchains. Proceedings of the Thirteenth EuroSys Conference, 1-15.

8. Benet, J. (2014). IPFS - Content addressed, versioned, P2P file system. arXiv preprint arXiv:1407.3561.

9. Sporny, M., Longley, D., & Chadwick, D. (2019). Verifiable Credentials Data Model 1.0. W3C Recommendation.

10. Ekblaw, A., Azaria, A., Halamka, J. D., & Lippman, A. (2016). A case study for blockchain in healthcare: "MedRec" prototype for electronic health records and medical research data. 2016 IEEE 2nd Open & Big Data Conference, 25-30.

11. Xu, X., Pautasso, C., Zhu, L., Gramoli, V., Ponomarev, A., Tran, A. B., & Chen, S. (2017). A taxonomy of blockchain-based systems for architecture design. 2017 IEEE International Conference on Software Architecture, 243-252.

12. Cachin, C., & Vukolić, M. (2017). Blockchain consensus protocols in the wild. arXiv preprint arXiv:1707.01873.

13. W3C DID Working Group. (2022). Decentralized Identifiers (DIDs) v1.0. W3C Recommendation.

14. Cheng, R., Zhang, F., Kos, J., He, W., Hynes, N., Fedorova, N., ... Song, D. (2019). Ekiden: A platform for confidentiality-preserving, trustworthy, and performant smart contracts. 2019 IEEE European Symposium on Security and Privacy, 185-200.

15. Ben-Sasson, E., Chiesa, A., Garman, C., Green, M., Miers, I., Tromer, E., & Virza, M. (2014). Zerocash: Decentralized anonymous payments from Bitcoin. 2014 IEEE Symposium on Security and Privacy, 459-474.

16. Dorri, A., Kanhere, S. S., Jurdak, R., & Gauravaram, P. (2017). Blockchain for IoT security and privacy: The case study of a smart home. 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), 618-623.

17. Kim, H., Park, J., Bennis, M., & Kim, S. L. (2020). Blockchained on-device federated learning. IEEE Communications Letters, 24(6), 1279-1283.

18. Zhang, P., White, J., Schmidt, D. C., Lenz, G., & RosenbloomS. T. (2018). FHIRChain: Applying blockchain to securely and scalably share clinical data. Computational and Structural Biotechnology Journal, 16, 267-278.

19. Nov, O. (2018). Blockchain meets IoT: An architecture for scalable access management in IoT. IEEE Internet of Things Journal, 5(2), 1184-1195.

20. Mengelkamp, E., Notheisen, B., Beer, C., Dauer, D., & Weinhardt, C. (2018). A blockchain-based smart grid: Towards sustainable local energy markets. Computer Science - Research and Development, 33(1-2), 207-214.

21. Shafagh, H., Burkhalter, L., Hithnawi, A., & Duquennoy, S. (2017). Towards blockchain-based auditable storage and sharing in IoT. Proceedings of the 7th ACM Conference on Data and Application Security and Privacy, 45-50.

22. Yeung, K. (2019). Regulation by blockchain: The emerging battle for supremacy between the code of law and code as law. Modern Law Review, 82(2), 207-239.