BLOCKSTAND

Accelerating Standards for On-Ledger and Off-Ledger Data

Introduction

Blockchain concepts and tecnology evolved over years and offered te world a technology solutions in decentralisation, information security and data transparency. However, with this growth comes increasing complexity in how data is managed, particularly the distinction between **on-ledger** (data stored directly on the blockchain) and **off-ledger** (data stored externally but linked to blockchain transactions). Managing both data types can be challenging at times because of the way it tackles security, privacy, interoperability, and legal enforceability. With the following report I will not dare to create form of a real standard, but more of an 'actionable document' that can drive further blockchain adoption and development in the field.

Objective of the Report

This report, **"Accelerating Standards for On-Ledger and Off-Ledger Data,"** aims to provide a structured framework for classifying and managing blockchain data. On a more granular level, it focuses on addressing how data could or in some cases should be split between on-ledger and off-ledger environments. My aims were to display the gentle balance of legality, security, privacy and cost-effectiveness that is often missed or unseen in the early stages of information technology systems using blockchain or DLT (distributed ledger). My other aim is to ensure the data integrity of the systems as a whole while to sacrificing their scalability and interoperability.

Importance of Standardisation

The need for standardisation in this area is especially critical within the **European Union's regulatory environment**, where compliance with frameworks such as GDPR, eIDAS, and MiCA is essential. Without formal standards, there is a risk of inconsistent practices across industries, which can hamper the scalability, trust, and adoption of blockchain technologies. Without overregulation but by establishing clear guidelines for managing on-ledger and off-ledger data, we can ensure that blockchain solutions remain secure, transparent, and compliant with evolving legal requirements.

SMEs: Beneficiaries of Clear Data Standards

While small and medium-sized enterprises (SMEs) may not directly grapple with the technical intricacies of on-ledger and off-ledger data management, they stand to gain significantly from the standardisation of these processes. SMEs represent a substantial portion of the economy, and their ability to adopt blockchain solutions hinges on the existence of clear, cost-effective, and scalable standards for data classification and management.

In industries like supply chain management, finance, and digital contracts—where many SMEs operate—the ability to efficiently manage which data resides on-chain versus off-chain will reduce operational complexity. Simplifying data interoperability and ensuring security will lower the barriers to blockchain adoption for SMEs, allowing them to leverage these technologies without needing extensive technical expertise.

By establishing robust guidelines for on-ledger and off-ledger data, this report helps ensure that blockchain is accessible to businesses of all sizes, ultimately fostering broader adoption and innovation across industries.

SMEs: Beneficiaries of Clear Data Standards

While small and medium-sized enterprises (SMEs) may not directly grapple with the technical intricacies of on-ledger and off-ledger data management, they stand to gain significantly from the standardisation of these processes. To some readers it might be new, that SMEs represent a substantial portion of the economy. Their ability to adopt the blockchain and the distributed ledger technology hinges on the existance of clear, cost-effective and scalable standards for data classification and management.

In industries like supply chain management, finance, and digital contracts—where many SMEs operate—the ability to efficiently manage which data resides on-chain versus off-chain will reduce operational complexity. During the report completion (September 2024) there are

still barriers for adoption to blockchain for SMEs and we would like to lower them simplifying data interoperability while of course ensuring security. Sme's should leverage the new technology without extensive technical expertise. A form of robust guidelines for on-ledger and off-ledger data will help ensuring that blockchain and distributed ledger technology are accessible to businesses of every sizes, ultimately fostering broader adoption across industries.

Scope of the Report

- **Data Classification**: Exploring how to determine which data should reside on-ledger versus off-ledger while taking in consideration factors such as transaction frequency, privacy concerns, and scalability.
- **Security**: Addressing the challenges of securing both on-ledger and off-ledger data, ensuring that off-ledger data remains verifiable and tamper-resistant, while maintaining the immutability of on-ledger records.
- **Interoperability**: Ensuring that systems can seamlessly exchange and synchronise on-ledger and off-ledger data across different blockchain networks and with legacy systems, promoting a more interconnected blockchain ecosystem.
- **Legal Enforceability:** Examining how on-ledger and off-ledger records affect the legal enforceability of smart contracts, and what legal frameworks are necessary to support blockchain-based agreements. This section would also address how off-ledger data, oracles, and hybrid models can ensure that smart contracts remain adaptable and compliant with evolving regulations.

Tackling these issues is not the most important part of the report but it will help me and if possible ather authors understand the foundation for developing more formal standards that will facilitate blockchain and DLT adoption into various business areas of life, with a tiny special focus on making it accessible by my beloved SMEs. Eventually this will ensure that the the technology supports businesses of all sizes, business areas, and continents in a compliant, secure and scalable way.

Two Important Notes:

1. While much of the discussion around blockchain standards focuses on **public blockchains** (the common term when we talk about blockchain), it’s important to recognise that **private networks** face similar challenges. In private or permissioned blockchains, which are often deployed by enterprises and organisations, the same issues of data classification, security, and scalability are present. In the report I am exploring solutions that apply across both **public blockchains** and **private networks**. It is rater intersting that both environments display a need for comprehensive standards.
2. I tried to abstain from using AI-helpers for standardatisation work, although I heavily rely on the technology for research and checking for correct pronounciation.

On-Ledger and Off-Ledger Data Classification

Blockchain systems are designed to manage transactions and store data in a decentralised and immutable manner. However, not all data can or should be stored directly on the blockchain (on-ledger). The very distinction between on-ledger and off-ledger data is important for understanding how technology perfors, drives cost efficiency, decent level of pricacy, scalability, and last but not least bounds data to some form of legal standards. Understanding the appropriate use cases for each type of data storage is key to developing standards that will guide further blockchain implementations.

Definition and Distinction

- **On-Ledger Data**: Stored directly on the blockchain, ensuring immutability and transparency. Examples: transaction records, smart contract execution, ownership certificates.
- **Off-Ledger Data**: Stored outside the blockchain, linked to on-ledger transactions to handle large data sets, sensitive information, or frequently changing data. Examples: large files, personal data, IoT sensor information.

On-Ledger Data: On-ledger data refers to information that is stored directly on the blockchain. This data benefits from the core features of blockchain: immutability, decentralisation, and transparency. Since blockchain records are tamper-proof once added to the ledger, on-ledger data is most often used for transactions, smart contract execution, and any critical data that must be publicly verifiable and permanent.

The **ISO/TS 23258:2022** standard, part of ISO/TC 307, provides guidelines for the architecture of **smart contracts** within blockchain systems, which are typically stored on-ledger due to the need for their transparency and immutability. This standard document is aiming to

help define how smart contracts interact with on-ledger data, making sure that key contract logic and execution are permanently stored on the blockchain.

Primary use cases for on-ledger data include:

- **Transaction Records**: Financial transfers, asset exchanges, or any other value transfer that requires an immutable, transparent record.
- **Smart Contracts**: Self-executing contracts that automatically enforce terms agreed upon by participants. The contract logic is recorded on the blockchain and executed by its network.
- **Tokenised Assets**: Ownership records, digital tokens, or NFTs that require transparency and trust, as they represent real or digital assets.

On-ledger data is stored in blocks that are part of a chain, ensuring that all network participants have access to a consistent, verifiable record of all transactions.

Off-Ledger Data: Off-ledger data, in contrast, is not stored directly on the blockchain but is often referenced or linked to on-ledger transactions. Off-ledger data storage often allows for greater flexibility and scalability, especially when handling large data sets, private information, or data that doesn't need to be publicly visible or immutable.

Standards such as **ISO/TR 23455:2019**, which outlines the interaction between **smart contracts** and **external data** (off-ledger), provide useful guidance. This standard explains how external data, often stored off-ledger, can interact with blockchain-based smart contracts, ensuring the integrity of off-ledger data while linking it to on-ledger transactions.

Primary use cases for off-ledger data include:

- **Large Data Files**: Documents, images, or video files that would be too costly or impractical to store on-chain.
- **Private Information**: Sensitive data such as personal identifiable information (PII), healthcare records, or internal business data, which requires secure storage but shouldn't be visible to all blockchain participants.
- **Reference Data for Transactions**: Data that is linked to an on-ledger transaction but does not need to be stored on the blockchain itself, such as digital signatures, proofs of ownership, or metadata related to assets.

Off-ledger data is often stored in traditional databases or decentralised storage solutions (like IPFS) and referenced by the blockchain through cryptographic methods, ensuring that it remains tamper-evident and accessible without being fully stored on the ledger.

Criteria for Classification

Deciding whether data should be stored on-ledger or off-ledger depends on several factors. The key considerations are **data size, privacy concerns, transaction frequency**, and **the need for transparency**.

The **ISO/TS 23635:2023 – Guidelines for Blockchain and DLT Governance**, also within ISO/TC 307, outlines governance strategies for blockchain ecosystems. These strategies can be applied when deciding whether data should be stored on or off the ledger, factoring in issues such as compliance, privacy, and data governance requirements.

- **Data Size**: Blockchain systems, particularly public ones, are not designed for handling large volumes of data due to the high cost of storage on the chain. Storing extensive data sets directly on the ledger can lead to scalability issues, higher transaction fees, and network congestion. Off-ledger storage solutions are better suited for large files, with only essential references or proofs being kept on the blockchain.
- **Privacy Concerns**: Blockchains are typically transparent, meaning all participants can see the data stored on-chain. This creates a conflict when handling sensitive or private information. GDPR and similar privacy regulations also restrict what data can be publicly visible. In such cases, off-ledger storage is preferred, with only encrypted hashes or references being stored on-chain, ensuring that personal data remains protected.
- **Transaction Frequency**: Frequently updated or ephemeral data is not well-suited for on-ledger storage due to blockchain’s immutable nature. Storing frequently changing data on-chain can result in high operational costs and inefficiency. Off-ledger storage is ideal for dynamic data, while the blockchain can reference a hash or state summary of this data for integrity.
- **Transparency and Verifiability**: On-ledger data should be chosen when public verifiability is required, such as in financial transactions or contract enforcement. In cases where transparency is less critical (e.g., private agreements or documents), off-ledger storage is sufficient, with the blockchain only referencing essential proofs (e.g., cryptographic hashes or digital signatures).

Approaches to Data Classification

Different approaches to classifying and linking on-ledger and off-ledger data have emerged based on the blockchain's specific use case. Some common methods include:

- **Hashing**: Large off-ledger data can be hashed, and only the hash is stored on the blockchain. This ensures that the integrity of the off-ledger data can be verified without exposing the data itself on-chain.
- **Digital Signatures**: Transactions and documents stored off-chain can be cryptographically signed, and the signature or reference is recorded on the blockchain to prove authenticity.
- **Metadata Storage**: Storing metadata on-chain, while keeping the bulk of the data off-chain, helps maintain the transparency and verifiability of key information without overburdening the blockchain.

Challenges and Considerations

The classification of on-ledger and off-ledger data is one of the main factors when approaching **cost**, **scalability**, and **performance**.

- **Cost Implications**: On-chain storage can be expensive, especially in public blockchains like Ethereum, where transaction fees (gas fees) increase with the amount of data being stored. Off-ledger storage significantly reduces costs, but requires careful management to ensure the data remains secure and verifiable.
- **Scalability**: Storing too much data on the blockchain can hinder scalability. As more data is added to the ledger, the time and resources needed for network nodes to process and validate transactions increase, potentially leading to slower transaction times and higher fees. Off-ledger storage reduces the burden on blockchain nodes and allows for the scaling of applications without compromising the network's performance.
- **Performance Considerations**: While off-ledger storage solutions alleviate the strain on blockchain networks, they also introduce complexity. Retrieving and verifying off-ledger data can increase latency, as the data is not stored directly within the blockchain. Balancing this trade-off is critical for ensuring both security and performance.
- **Balancing Transparency and Privacy**: A key challenge when choosing between on-ledger and off-ledger storage is finding the right balance between the need for transparency and the obligation to protect private data. While the blockchain ensures data is transparent and immutable, privacy concerns necessitate keeping certain data off-chain. The use of techniques like zero-knowledge proofs and encrypted hashes allows for privacy-preserving verification without compromising transparency.

Cost Implication Examples

Let's expand on the previous section by adding examples of **cost comparisons for blockchain transactions** and **on-chain record-keeping requirements** for different blockchain platforms and networks. This will provide practical insights into the **minimum and maximum costs**, and the **data management approaches** used by various blockchains.

Cost of Blockchain Transactions on Public Networks: Minimum and Maximum

The cost of blockchain transactions varies greatly depending on the blockchain's consensus mechanism, network congestion, and the complexity of the data or transactions involved. Here's a range of examples from low-cost to high-cost blockchains:

1. **Nano (Minimum Cost - Close to Free)**
   - **Transaction Cost**: Nano has a fee-less transaction model. This is possible because Nano uses a block-lattice structure and a **Proof-of-Work (PoW)** anti-spam system, which requires only a minimal computational effort for each transaction.
   - **Use Case**: Nano is best suited for small, instant value transfers. It offers low-latency transaction processing, ideal for payments and micropayments.
2. **Solana (Low Cost)**
   - **Transaction Cost**: On **Solana**, the average transaction cost is around **$0.00025 USD**. Solana's high throughput (up to 65,000 transactions per second) makes it one of the most cost-effective blockchain platforms for high-frequency transactions.
   - **Use Case**: Solana is often used for DeFi applications, payments, and NFTs, where fast processing times and low costs are essential.
3. **Tezos (Low to Moderate Cost)**

- **Transaction Cost**: Tezos charges around **$0.01 to $0.10 USD** per transaction, depending on network conditions and transaction complexity.
- **Use Case**: Tezos is widely used for smart contracts and decentralised applications (dApps). Its proof-of-stake (PoS) consensus mechanism allows for lower transaction fees compared to PoW systems.

4. **Polygon (Moderate Cost)**
   - **Transaction Cost**: Polygon, an Ethereum Layer 2 scaling solution, has transaction costs that vary but typically range between **$0.01 to $0.10 USD**. This is much lower than the fees on Ethereum's main network.
   - **Use Case**: Polygon is often used to scale dApps, particularly for DeFi and NFT platforms, where lower transaction costs are essential for a better user experience.
5. **Ethereum (High Cost - Varies Based on Network Congestion)**
   - **Transaction Cost**: Costs on Ethereum vary significantly, but in high-demand periods, transactions can cost anywhere from **$5 to $200 USD**. Ethereum uses a gas model to calculate transaction fees based on computation and network usage.
   - **Use Case**: Ethereum is a preferred platform for decentralised finance (DeFi), smart contracts, and NFT marketplaces. Its transaction fees have made it less practical for small transactions but suitable for more complex, high-value use cases.
6. **Bitcoin (Highest Cost During Peak Usage)**
   - **Transaction Cost**: Bitcoin transaction fees fluctuate based on network congestion, with average fees typically around **$1 to $10 USD**, but they have peaked at over **$60 USD** during periods of extreme congestion, such as during bull markets.
   - **Use Case**: Bitcoin is primarily used for transferring high-value assets or serving as a store of value. Its high fees make it less ideal for small, frequent transactions.

On-Chain Record Keeping on Public Networks: Minimum and Maximum Data Requirements Across Platforms

Different blockchain platforms have varying requirements for **on-chain record-keeping**, depending on their design, consensus mechanisms, and intended use cases. Here's a breakdown of minimum and maximum storage limits across a range of platforms:

1. **Bitcoin**
   - **On-Chain Record Keeping**: Bitcoin stores basic transaction data (sender, receiver, amount, timestamp) on-chain. Each block can store **up to 1 MB** of transaction data.
   - **Use Case**: Primarily used for value transfers, with limited capability for storing additional data beyond transactions.
2. **Ethereum**
   - **On-Chain Record Keeping**: Ethereum stores not only transaction data but also smart contracts and their execution data. A block can store **up to 30 million gas units**, which translates to approximately **50 to 80 kB** of actual data, depending on the operations.
   - **Use Case**: Ethereum supports more complex dApps, smart contracts, and NFTs, where not just transaction data but programmatic logic is stored on-chain.
3. **Hyperledger Fabric**
   - **On-Chain Record Keeping**: Hyperledger Fabric allows for both on-chain and off-chain storage, with a focus on **privacy** and **modular architecture**. While transactional metadata is stored on-chain, large datasets or sensitive information are typically stored off-chain.
   - **Use Case**: Hyperledger Fabric is used in enterprise environments where certain data (such as private contracts or business data) must remain off-chain due to privacy concerns or performance limitations.
4. **Stellar**
   - **On-Chain Record Keeping**: Stellar stores basic transaction data on-chain, primarily focused on cross-border payments. A single operation on Stellar consumes around **100 bytes**, and blocks (or ledgers) are confirmed every 5 seconds.
   - **Use Case**: Stellar is used for cross-border payments, asset issuance, and low-cost remittances, where only minimal transaction data is kept on-chain.
5. **Polkadot**
   - **On-Chain Record Keeping**: Polkadot enables interoperability between different blockchains and allows parachains to store custom data on-chain, but the main relay chain only stores **transaction validation** data. Parachains can vary in their storage requirements, offering flexibility for developers.

- **Use Case**: Polkadot is suited for blockchain interoperability, where different chains may store varying amounts of data on-chain depending on their specific use cases.

6. **Arweave (Maximum Storage)**:
   - **On-Chain Record Keeping**: Arweave is a blockchain protocol specifically designed for permanent data storage. It supports **up to 1 MB** of on-chain data storage per transaction and allows larger files (up to several gigabytes) to be stored via linked systems.
   - **Use Case**: Arweave is used for archival purposes, where large files, documents, and records need to be stored immutably on-chain or via linked off-chain storage.

Conclusion: Minimum and Maximum Requirements

- **Minimum Costs**: Fee-less or near-fee-less transactions are possible on **Nano** and **Solana**, while **Tezos** and **Polygon** offer moderate fees ranging from $0.01 to $0.10 USD.
- **Maximum Costs**: At the high end, platforms like **Ethereum** and **Bitcoin** can see transaction fees rise significantly, particularly during periods of congestion, with costs ranging from **$5 to $200+ USD**.
- **On-Chain Storage**:
  - **Minimal Storage**: Bitcoin (1 MB per block) and Stellar (minimal transaction data) focus on storing the essentials, while Ethereum (30 million gas units) offers slightly more data storage for transactions and smart contracts.
  - **Maximal Storage**: Platforms like **Hyperledger Fabric** and **Arweave** allow for hybrid and large-scale on-chain or linked data storage, supporting complex applications like enterprise use cases and permanent data archives.

The abovelisted examples illustrate the **cost and scalability trade-offs** involved in choosing between on-chain and off-chain storage, along with the different data requirements across public blockchain platforms.

Cost Implications in Private Permissioned Networks

While cost implications are often a significant concern in **public blockchains**—where transaction fees can fluctuate depending on network demand and congestion—**private permissioned networks** operate under a different model. However, cost concerns still exist, albeit in a different context.

In **public blockchains**, transaction fees (e.g., gas fees on Ethereum or transaction fees on Bitcoin) are required to incentivise network validators. Conversely, **private permissioned networks** don't rely on open, decentralised consensus mechanisms that require such fees. Instead, these networks are typically maintained by a consortium or a central entity, which can control access and operation costs more directly.

Cost Drivers in Private Networks

In private permissioned networks, the costs are not tied to market-driven fees but rather to **infrastructure**, **management**, and **maintenance**. Some of the key cost factors include:

- **Infrastructure Costs**: Running and maintaining the nodes that validate transactions is typically managed by the entities participating in the network. This involves **hardware, software, and cloud service costs** associated with maintaining a distributed ledger system, even if there are fewer nodes than in public blockchains.
- **Operational Costs**: Managing data on a private blockchain, especially in terms of **on-ledger** and **off-ledger** classification, involves operational costs. This includes deciding how much data should be stored on-chain (which might impact performance and system scalability) versus off-chain storage, which could introduce complexity but may reduce the burden on the network infrastructure.
- **Consensus Costs**: Unlike public blockchains, where consensus mechanisms like **Proof-of-Work** (PoW) and **Proof-of-Stake** (PoS) are resource-intensive, private blockchains often use more efficient mechanisms like **Byzantine Fault Tolerance (BFT)** or **Proof of Authority (PoA)**. These require less computational power, but maintaining network security and data integrity still carries an associated cost.

Examples of Private Blockchains and Their Cost Considerations

1. **Healthcare Blockchains (Electronic Health Records - EHRs)**:

- **On-Ledger**: Patient metadata, such as cryptographic proofs of record authenticity, patient consent records, and treatment history, may be stored on the blockchain to ensure data integrity and allow for secure audit trails.
- **Off-Ledger**: Due to privacy concerns and the large size of medical records (images, scans, full health histories), these are stored off-chain in secure, private storage systems, with only references or hashes stored on-chain.
- **Cost Impact**: In this scenario, the cost is driven by the need to balance privacy with transparency, ensuring that patient information remains secure while maintaining verifiability. Storing only essential verification data on the blockchain helps reduce infrastructure costs.

2. **Financial Services (Trade Finance & Digital Currencies)**:
   - **On-Ledger**: Key transaction data, contract terms, and proof of ownership are stored on-chain to ensure immutability and provide a trusted record of trades.
   - **Off-Ledger**: Supporting documents (e.g., invoices, contracts, bills of lading) that are too large to store on-chain are stored off-ledger. These documents are linked to the blockchain through hashes or other cryptographic references.
   - **Cost Impact**: The infrastructure and operational costs can be significant, as financial institutions require **real-time processing** of transactions with stringent **compliance** requirements. The efficient use of off-ledger storage reduces the burden on the blockchain, keeping costs in check while ensuring compliance with financial regulations.
3. **Supply Chain Blockchains**:
   - **On-Ledger**: Key information such as timestamps, proof of authenticity, and ownership changes for products in the supply chain are stored on-chain.
   - **Off-Ledger**: Detailed product descriptions, manufacturing data, and sensitive contractual information are stored off-ledger to avoid bloating the blockchain.
   - **Cost Impact**: Storing detailed supply chain data off-chain helps reduce costs by ensuring that only critical information is placed on the blockchain, maintaining both transparency and scalability. Companies avoid the high costs of blockchain storage while ensuring supply chain visibility.
4. **Energy Trading (Smart Grids and Decentralised Energy Markets)**:
   - **On-Ledger**: Metering data, consumption records, and energy transaction records are stored on-chain to ensure verifiability and facilitate peer-to-peer energy trading.
   - **Off-Ledger**: Large datasets, such as full energy consumption profiles and real-time monitoring data, are stored off-ledger. These datasets are referenced on the blockchain through digital signatures or hashes.
   - **Cost Impact**: Energy networks generate vast amounts of data that are impractical to store directly on the blockchain. Storing only essential data on-chain reduces the overall costs and maintains the efficiency and speed of the energy trading platform.
5. **Logistics & Transportation (Fleet Management)**:
   - **On-Ledger**: Tracking data for high-value goods and proof of delivery are stored on-chain, ensuring verifiable audit trails and transparency.
   - **Off-Ledger**: Detailed fleet management data (e.g., GPS tracking, driver logs, fuel usage, and maintenance records) are stored off-chain due to the volume of data.
   - **Cost Impact**: In logistics, the scalability and efficiency of the blockchain system are critical. Storing only tracking information on-chain reduces the storage load, making it more cost-effective to run the network, while off-chain solutions handle the bulk of the data.
6. **Intellectual Property (IP) and Content Licensing**:
   - **On-Ledger**: Proof of ownership, licensing agreements, and time-stamped records for content creation (e.g., music, videos, digital artwork) are stored on-chain.
   - **Off-Ledger**: Full content files (audio, video, images) are stored off-chain to avoid large data costs and maintain privacy.
   - **Cost Impact**: In industries such as content licensing, on-chain verification of ownership and licensing is crucial for dispute resolution. However, storing the actual media on the blockchain would be prohibitively expensive. Therefore, blockchain is only used for verification, helping content creators maintain rights without incurring excessive costs.

Though private permissioned networks avoid the volatile transaction fees of public blockchains, they still face significant infrastructure, management, and operational costs. The decision to store data on-ledger or off-ledger directly affects these costs, with scalability and data management being key factors. Efficient data classification, especially in private networks, helps organisations balance the need for secure and verifiable data without overburdening their infrastructure.

These examples from healthcare, finance, supply chains, energy markets, logistics, and intellectual property show that while private networks face different cost structures than public blockchains, they must still carefully manage their on-ledger and off-ledger data to ensure scalability and cost-effectiveness.

Scalability and Performance Examples

Let's expand on the scalability and performance challenges by examining specific examples from both **private permissioned** and **public blockchain networks**. My aim is to examplify how different platforms handle **on-ledger and off-ledger data** management and thus the resulting impact on transaction throughput, data storage, and overall scalability. I hope that by exploring the solutions employed—such as **Layer 2** and **Layer 3** protocols—we can gain practical insights into how networks optimise performance, manage large datasets, and balance transparency and security. These examples aim to highlight the varying approaches to improving network efficiency while addressing the challenges associated with on-ledger and off-ledger data in different blockchain ecosystems.

Scalability and Performance Considerations for Public Networks

The blockchain salesman is telling us that Public networks, such as **Bitcoin**, **Ethereum**, and others, have gained widespread adoption due to their decentralised, transparent, and secure nature. What the salesman is not telling you is that these networks are followed by a persistent challenges (for more than 15 years) in **scalability** - in other words, the ability to handle a growing number of transactions efficiently without compromising speed, cost, or security. Blockchain users often meet performance issues when the network becomes congested, leading to increased transaction fees and slower confirmation times. As a result, many projects have begun exploring **Layer 2** and even **Layer 3** solutions to address these challenges.

In addition to the issues of transaction throughput and performance, the distinction between **on-ledger** and **off-ledger data** plays a critical role in the scalability of public networks. Deciding which data should be stored directly on the blockchain (on-ledger) and which should remain off-chain (off-ledger) can significantly impact network performance and overall efficiency.

Scalability Challenges in Public Networks

Public blockchains like **Ethereum** and **Bitcoin** are constrained by their **Layer 1** architecture, which has inherent limitations on how many transactions can be processed per second (TPS). For instance:

- **Bitcoin**: Can process around **7 TPS** due to its limited block size and relatively long block time.
- **Ethereum**: Currently handles around **15-30 TPS**, but transaction times and fees often spike during periods of high network demand, as seen during the NFT boom and DeFi activity surges.

These limitations are exacerbated when **large amounts of data** are stored directly on the blockchain. **On-ledger data** includes transactions, smart contracts, and essential records that require transparency and immutability. However, as the amount of on-ledger data grows, it increases the size of the blockchain, making it more resource-intensive for nodes to validate and store the data.

On-Ledger vs. Off-Ledger Data and Scalability

The distinction between on-ledger and off-ledger data becomes crucial when discussing scalability:

- **On-Ledger Data**: On-ledger data refers to information that is stored directly on the blockchain, benefiting from immutability, decentralisation, and security. This includes transaction records, smart contracts, and tokenised assets. However, storing large amounts of on-ledger data increases the size of the blockchain, requiring more storage and processing power from nodes, which can slow down the network and raise transaction costs.
- **Off-Ledger Data**: Off-ledger data, in contrast, is stored outside the blockchain but referenced or linked to on-ledger transactions through cryptographic methods (e.g., hashes or proofs). This approach reduces the burden on the blockchain by keeping large files (e.g., images, documents, or other non-essential data) off-chain. Off-ledger solutions, such as **IPFS** or traditional cloud storage, help maintain scalability by allowing the blockchain to focus on transaction verification and critical data while still ensuring data integrity through off-chain references.

Layer 2 Solutions: Enhancing Scalability

To mitigate the scalability issues caused by storing large amounts of on-ledger data, **Layer 2 solutions** have emerged. These solutions aim to handle transactions and data processing off-chain (outside the main blockchain) while leveraging the security of the main chain.

1. **Lightning Network (Bitcoin)**:
   - **Description**: The Lightning Network allows Bitcoin transactions to occur off-chain through a network of payment channels. Only the opening and closing of these channels are recorded on the Bitcoin blockchain, reducing the amount of on-ledger data.
   - **Impact**: This significantly improves Bitcoin's scalability by keeping many microtransactions off-ledger while ensuring that essential transaction finality remains on the main chain.
2. **Optimistic Rollups and ZK-Rollups (Ethereum)**:
   - **Description**: These Layer 2 solutions bundle multiple transactions off-chain and submit a single proof or summary to Ethereum's mainnet. By moving bulk transactions and computationally heavy processes off-ledger, they reduce on-chain congestion.
   - **Impact**: This increases Ethereum's capacity for transactions and data storage while ensuring that only the most critical data remains on the main chain, optimising on-ledger storage for essential records.
3. **Polygon (Layer 2 for Ethereum)**:
   - **Description**: Polygon is a Layer 2 scaling solution that processes transactions on sidechains before submitting the results to the Ethereum mainnet. It also allows for the management of large off-ledger datasets while maintaining security through cryptographic references.
   - **Impact**: Polygon helps keep unnecessary data off the Ethereum mainnet, ensuring that on-ledger storage is reserved for essential transaction records and reducing network congestion.

Layer 3 Solutions: Extending Beyond Layer 2

While **Layer 2 solutions** have significantly improved scalability, some projects are now exploring even further - **Layer 3 solutions**, which build on Layer 2 to handle even more complex use cases, including advanced data storage models for **on-ledger** and **off-ledger** data. Layer 3 networks may provide even more efficient handling of large datasets, specialised applications, and cross-chain communication.

For example:

- **Immutable X**: A Layer 2 protocol built on **ZK-Rollups** for scaling Ethereum's NFT market, which could evolve into a Layer 3 for **specialised applications** requiring extensive off-ledger data management for digital assets.
- **StarkEx**: A Layer 2 solution using **zero-knowledge proofs** to scale Ethereum. In the future, this could evolve into a Layer 3 solution that handles large datasets off-ledger while maintaining verification and integrity through Layer 2 rollups.

Remaining Challenges and Future Needs

Despite the progress made by **Layer 2** and emerging **Layer 3** technologies, further work is required to handle the challenges of **on-ledger** and **off-ledger** data, particularly as blockchain networks scale. Some of the ongoing challenges include:

1. **Efficient Data Storage**: As more data is stored on public blockchains, the size of the ledger continues to grow, leading to increased resource requirements for nodes. Off-ledger storage helps mitigate this issue, but the need for secure and verifiable integration between on-chain and off-chain data remains a key challenge.
2. **Data Retrieval**: Off-ledger solutions like **IPFS** or decentralised storage systems can reduce the burden on the blockchain, but they must ensure quick and reliable data retrieval, especially for time-sensitive applications like DeFi or supply chain management.
3. **Interoperability**: Ensuring seamless interaction between **on-ledger** and **off-ledger** data across different Layer 2 and Layer 3 solutions is still an evolving area. Standards for interoperability are needed to allow different networks to securely share off-chain data without undermining the integrity of the on-chain records.
4. **User Experience**: While Layer 2 and Layer 3 solutions improve scalability, interacting with these multiple layers and off-chain storage can create a confusing user experience. Simplifying the process of using off-ledger storage while ensuring security and ease of use will be essential for broader adoption.

Scalability and Performance Considerations for Private Networks

Private permissioned **DLT** networks, such as **Corda**, **Hyperledger Fabric**, and **Hyperledger Besu**, are designed to offer more control, privacy, and governance than public DLTs. by design private networks should mean less broadcasting and propagading ledger updates, but alas, no- networks still face significant scalability and performance challenges, particularly when it comes to managing **on-ledger** and **off-ledger data**. I do agree that some progress has been made in addressing these issues (I have the chance to supersede how Corda operates and gain real-life stats from projects 'in production', but many problems remain unresolved, particularly around data management, network congestion, and transaction throughput.

On-Ledger vs. Off-Ledger Data in Private Networks

In private networks, the classification of **on-ledger** and **off-ledger data** is critical to optimising performance and ensuring scalability. These systems often store transactional metadata and cryptographic proofs on-chain (on-ledger) while moving larger data, such as documents, contracts, or sensitive information, off-chain (off-ledger). This approach aims to reduce the load on the distributed ledger, enabling faster processing times and improved throughput. However, this method presents unique challenges:

- **On-Ledger Data**: Storing too much data on the ledger can cause performance bottlenecks, particularly in networks with multiple participants. While on-ledger data provides transparency and immutability, it can lead to increased ledger size, making it harder for nodes to keep up with validation and consensus.
- **Off-Ledger Data**: Off-ledger data storage solutions are often used to handle large files and sensitive information that do not need to be stored directly on the ledger. However, this introduces **complexity in data retrieval** and **verification**, as the integrity of off-ledger data must be maintained without sacrificing security.

Scalability Challenges in Private Networks

Despite their focus on performance and control, private **DLT** networks still face significant scalability issues when handling **on-ledger** and **off-ledger data**. A few notable examples highlight these challenges:

1. **Corda (Versions 4 and 5)**
   - **Scalability Issues**: Corda has encountered substantial scalability problems in **version 4**, primarily due to its reliance on point-to-point communication between nodes and its use of **on-ledger** data to maintain consensus. The network struggles with increasing transaction volume and the need to process large datasets.
   - **Off-Ledger Data**: While Corda allows for off-ledger storage of non-critical data, the integration between off-chain and on-chain information is complex and lacks sufficient optimisation, further contributing to its scalability challenges. **Version 5**, which is expected to address these issues, is still under development and is not yet capable of solving them fully.
2. **Hyperledger Fabric**
   - **Scalability Concerns**: Hyperledger Fabric also faces scalability hurdles, particularly when managing **on-ledger** transactions. The network allows organisations to define what data is stored on-chain, but the increased use of **smart contracts** and complex business logic has led to performance degradation as transaction volume grows.
   - **Off-Ledger Data**: Hyperledger Fabric uses **private data collections** for off-ledger storage, but maintaining the security and integrity of this off-chain data while linking it to on-ledger proofs is challenging. Fabric's endorsement policies create additional overhead, which affects throughput, especially in large networks.
3. **Hyperledger Besu**
   - **Scalability Issues**: As an Ethereum client, Hyperledger Besu inherits some of the same scalability issues seen in Ethereum, such as slow transaction throughput when using **on-ledger** data storage. Besu's use in private settings has not fully solved these problems, and its capacity for handling large amounts of data on-chain remains a bottleneck.
   - **Off-Ledger Data**: While Besu supports the use of **private transactions** and off-ledger data handling, ensuring that these remain consistent with on-chain activity without causing delays or increasing complexity remains a challenge. Reports suggest that scaling large networks with complex data flows in Besu is still a work in progress.

Current Solutions and Future Needs

Though private networks like Corda, Hyperledger Fabric, and Besu have made strides in handling **on-ledger** and **off-ledger data**, scalability challenges persist. The key issues include:

1. **Data Integration**: Integrating **off-ledger** data with **on-ledger** proofs remains difficult. Networks need to ensure that off-ledger data is easily retrievable, secure, and verifiable without causing delays or additional costs.
2. **Performance Bottlenecks**: As networks grow, performance bottlenecks occur when too much **on-ledger** data is stored, leading to longer transaction times and higher resource consumption. Efficient **off-ledger** data management is essential, but it requires better coordination and faster data retrieval solutions.
3. **Network Congestion**: As more participants join private networks, the transaction validation process slows down, particularly when managing large datasets. This leads to increased latency and reduced throughput, making the network less scalable.

While private networks steadily offer more control and customisation than public DLTs, they still face significant challenges when it comes to **scalability** and the management of **on-ledger** and **off-ledger data**. As seen with **Corda's ongoing scalability issues** in versions 4 and 5, and the similar difficulties reported by **Hyperledger Fabric** and **Besu**, the problem remains far from solved. More engineering work is required to optimise these networks for large-scale data management while maintaining the security and integrity of both **on-ledger** and **off-ledger data**. Future versions and updates to these platforms must address these fundamental issues to fully unlock their potential for enterprise use.

Enterprise Hesitation: Abandoned and Stalled DLT Projects Due to Unmet Performance Needs

Reports from enterprises indicate that while **private permissioned DLTs** offer security and control, many projects have faced significant scalability and performance challenges related to the handling of **on-ledger** and **off-ledger data**. The inability to process large transaction volumes, handle complex data sets, and maintain efficient performance has caused hesitation or rejection of these solutions. As a result, many enterprises are either limiting their DLT deployments, awaiting further technological improvements, or seeking alternative architectures.

1. Corda's Scalability Issues

- **Reports and Feedback**: Enterprises using Corda (especially **Corda 4**) have reported significant scalability issues, particularly due to its point-to-point communication model. As transaction volume increases, the network faces bottlenecks. Managing **on-ledger** and **off-ledger data** has also proven complex, leading to performance degradation when too much data is stored on-ledger. This has caused some enterprises to reconsider Corda for large-scale projects, as **version 5** has yet to fully resolve these issues.
- **Enterprise Response**: Some enterprises have opted to pause or scale down their Corda projects, awaiting future updates to address these challenges or looking for alternative solutions.

2. Hyperledger Fabric Scalability Problems

- **Reports and Feedback**: Hyperledger Fabric, widely used by enterprises for permissioned DLT networks, has been criticised for its performance in large-scale deployments. One of the primary issues arises when managing **on-ledger transactions** that involve complex business logic and smart contracts. Enterprises have reported that the platform struggles with high transaction volumes and the need to process large **off-ledger data** sets, particularly when integrating external data sources.
- **Enterprise Response**: Some businesses in industries such as supply chain, healthcare, and finance have chosen to limit the scope of their Fabric deployments or use alternative architectures due to performance concerns.

3. Hyperledger Besu (Ethereum-based):

- **Reports and Feedback**: While **Hyperledger Besu** brings Ethereum functionality into the private DLT space, it inherits many of the scalability issues from the Ethereum network itself. Enterprises have noted that managing **on-ledger data** becomes problematic at scale, especially as transaction throughput remains low compared to other solutions. Handling **off-ledger data** has been difficult due to performance issues, particularly in large networks with many participants.
- **Enterprise Response**: Some organisations have moved away from Besu or postponed implementation due to concerns over scalability and the need for higher performance in enterprise settings.

4. IBM and Maersk (TradeLens Project)

- **Reports and Feedback**: The **TradeLens** project, developed by **IBM** and **Maersk** using **Hyperledger Fabric**, faced issues with scaling as the network expanded globally. The challenge of managing **on-ledger data** (for verifiable transactions) while handling large datasets off-ledger created performance bottlenecks, particularly when multiple participants needed to retrieve data efficiently. Though the project was one of the highest-profile blockchain applications in trade, these scalability issues contributed to its eventual winding down in late 2022.
- **Enterprise Response**: TradeLens was ultimately shut down, and similar reports have emerged from other large-scale blockchain consortia, where scalability and performance concerns led to project halts or cancellations.

5. General Enterprise Hesitation

- **Scalability vs. Performance Trade-offs**: Many reports and case studies indicate that enterprises hesitate to adopt **private permissioned DLTs** because of the **trade-offs** between achieving scalability and maintaining performance. In permissioned environments where trust is partially decentralised (rather than fully decentralised like public DLTs), performance issues arise when there are too many validators or when **on-ledger data** is used inefficiently. The complexity of managing large datasets and ensuring that off-ledger data remains verifiable without burdening the network has been a sticking point for many enterprises.
- **Reports**: Publications by the **World Economic Forum (WEF)** and **Gartner** have repeatedly highlighted that while **DLTs** show promise, scalability issues remain a **major barrier to enterprise adoption**, especially when large datasets are involved. Enterprises in industries such as finance, healthcare, and logistics have cited scalability as a core reason for postponing DLT projects.

Balancing Transparency and Privacy: On-Ledger vs. Off-Ledger Data Storage

In both **public** and **private distributed ledger technologies (DLTs)**, a fundamental tension exists between the need for **transparency** and the necessity for **privacy**. The decision to store data **on-ledger** (directly on the blockchain) or **off-ledger** (in external systems but referenced on the blockchain) often revolves around finding the right balance between these two requirements. In both contexts, managing this balance is critical to ensuring that DLTs are scalable, secure, and compliant with regulatory requirements, while also meeting the practical needs of the users.

Transparency vs. Privacy: A Persistent Challenge

1. **Public Networks**:
   - **Transparency** is a defining characteristic of public blockchains (DLTs), such as **Bitcoin** and **Ethereum**, where all transactions and data stored on-ledger are visible to every participant in the network. This transparency ensures **trust**, **auditability**, and **immutability**, making public DLTs valuable for use cases where verification and public accountability are necessary.
   - However, public transparency conflicts with the need for **privacy**, particularly when dealing with sensitive information such as personal data (e.g., healthcare records, financial details) or intellectual property. For instance, the **GDPR** (General Data Protection Regulation) in the EU mandates strict controls over how personal data is handled, which is difficult to reconcile with the immutability of on-ledger data.
2. **Private Networks**:
   - In **private DLTs** like **Corda** and **Hyperledger Fabric**, **privacy** is often prioritised because these networks are used by enterprises that need to protect sensitive business data or comply with industry regulations. While participants in private networks have more control over who can view on-ledger data, transparency is often reduced to only those with appropriate permissions, leading to potential trust issues when it comes to auditability or third-party verification.
   - Even in private networks, companies are reluctant to store sensitive data on-ledger for fear of exposing critical business information, trade secrets, or personally identifiable information (PII) to other participants in the network.

Why Do We Need Off-Ledger Data?

In both public and private DLTs, **off-ledger data storage** has become a necessity to address privacy concerns. Off-ledger storage allows for larger datasets, sensitive personal information, or proprietary business data to be stored externally, while still linking this data to the on-ledger transaction for **integrity and verifiability**. There are several key reasons why **off-ledger storage** is essential:

- **Data Privacy**: Keeping certain information off-chain ensures that sensitive data remains private, reducing the risk of exposing personal or business-critical information to unauthorized parties. This is especially important in industries like healthcare, finance, and government, where regulations like **GDPR** or **HIPAA** mandate strict privacy controls.
- **Scalability**: Off-loading large datasets to external storage prevents blockchain bloat, helping maintain the efficiency of the DLT network. Storing data off-chain can improve transaction throughput and reduce network congestion, allowing blockchains to scale more effectively.
- **Regulatory Compliance**: By keeping personal or sensitive information off-ledger and only storing cryptographic proofs or references on-ledger, organisations can comply with privacy laws while still benefiting from the transparency and immutability that DLTs provide.

Is There a Way to Keep Data Private While Storing It On-Ledger?

While off-ledger storage is often the default solution for managing sensitive data, there is growing interest and research into methods that allow sensitive data to be stored **on-ledger** while still maintaining privacy. Some key solutions and research areas include:

1. **Zero-Knowledge Proofs (ZKPs)**:
   - **ZKPs** are cryptographic techniques that allow one party to prove the validity of a transaction or data without revealing the underlying information. This allows sensitive data to remain private while still being stored or referenced on the ledger. ZKPs are already being used in some public networks like **Zcash**, and there is ongoing research to expand their application to more use cases in both public and private DLTs.
2. **Homomorphic Encryption**:
   - This encryption method allows computations to be performed on encrypted data without decrypting it. This means that sensitive data could be stored on-ledger in an encrypted form, allowing the network to process and verify the data without revealing its contents. While promising, **homomorphic encryption** is still in its early stages and presents challenges in terms of computation overhead and performance.
3. **Confidential Transactions**:
   - **Confidential transactions** are used to hide the amounts being transacted on a blockchain while still allowing the network to verify the transaction's validity. This technique has been used in cryptocurrencies like **Monero** and **Bitcoin's Liquid Network**, providing more privacy while maintaining on-ledger storage.
4. **Private Transactions and Channels (Corda, Hyperledger)**:
   - Private permissioned DLTs such as **Corda** and **Hyperledger Fabric** offer **private transaction** capabilities, allowing sensitive data to **only be visible** to authorised parties within a transaction, while cryptographic proofs or hashes are stored on the ledger for verification. These solutions are effective in closed networks but may be less feasible for public or semi-public DLTs.

Research and Future Directions

We see the need to balance **transparency** and **privacy** to spurr significant research in the DLT space. Several ongoing initiatives focus on improving the ability to handle private data on-ledger while ensuring scalability and performance:

- **ZK-Rollups and Optimistic Rollups** (Ethereum Layer 2 solutions): These scaling solutions combine **off-chain computation** with **on-chain verification**, enabling the blockchain to verify batches of transactions without revealing the private details. This approach is promising for reducing on-chain congestion while maintaining privacy for the underlying data.
- **Decentralized Identity (DID)**: Research in the field of **decentralised identities** aims to give individuals control over their personal data while ensuring that only the minimal amount of data is stored or verified on-ledger. This could provide a framework for enabling privacy-focused applications on public DLTs without sacrificing transparency.
- **Data Protection Standards**: Efforts like **ISO/TC 307** are working to standardise privacy-enhancing technologies in DLTs, helping to formalise the use of privacy-preserving techniques like **ZKPs**, **trusted execution environments (TEEs)**, and **private channels**.

Security Considerations for On-Ledger and Off-Ledger Data

As noted in the on-ledger/off-ledger classification, on-ledger data benefits from blockchain's inherent immutability, while off-ledger data introduces new security challenges, such as ensuring data verifiability through cryptographic proofs.

In **distributed ledger technologies (DLTs)**, security considerations for **on-ledger** and **off-ledger** data are critical. While **on-ledger data** benefits from transparency and immutability, it is exposed to certain risks, particularly in **public networks**. **Off-ledger data**, though its greater privacy and scalability, introduces several new challenges around security, including risks of tampering, loss of control, and how to ensure verifiability. Engineers find it challenging to link the **on-ledger** and **off-ledger data** securely while maintain the performance and integrity of the entire system.

Security Considerations for On-Ledger Data

1. **Cryptographic Integrity and Immutability**:
    - **Cryptographic integrity** ensures that once data is stored on the ledger, any attempt to alter it is immediately detectable. This is achieved through **hashing** and the use of **Merkle Trees**, where data is broken into smaller chunks and cryptographically hashed. The hashes are then combined in a tree structure, culminating in a **Merkle root**, which is stored on the ledger to represent the entire data set. Any change in the underlying data would result in a different hash, compromising the chain of integrity.
    - **Immutability** is enforced through **consensus mechanisms** such as **Proof of Work (PoW)** or **Proof of Stake (PoS)**, ensuring that once data is validated and recorded on the ledger, it cannot be changed. In such a scenario on-ledger data is ideal for applications requiring trust and verifiable audit trails.
2. **Exposure of On-Ledger Data and Associated Risks**:
    - **Public exposure** is a critical risk when storing data directly on a public ledger. In public DLTs, on-ledger data is visible to every participant, leading to potential breaches of privacy, especially if sensitive information is inadvertently stored on-chain. Even if cryptographic techniques like **hashing** are used, **metadata** and **transaction details** (such as sender and receiver addresses or transaction values) can expose patterns that could compromise privacy.
    - **Linkability and surveillance**: Even if data is hashed, an observer could track transaction patterns to deduce information about participants, leading to de-anonymisation. We can highlight that as a significant concern in privacy-sensitive industries like healthcare and finance, where **personally identifiable information (PII)** is involved.
    - **Legal and compliance risks**: Storing immutable on-ledger data can create problems with regulations such as **GDPR**, which require data deletion capabilities, including the "right to be forgotten." Once sensitive data is recorded on the blockchain, it becomes difficult—if not impossible—to remove, creating potential legal challenges.

Security for Off-Ledger Data

Given the risks associated with **on-ledger data**, particularly around public exposure, many enterprises store sensitive information **off-ledger**. However, off-ledger storage introduces its own set of security risks and considerations.

1. **Risks of Storing Data Off-Ledger**:
    - **Loss of Control**: Once data is stored off-ledger, especially in external databases or decentralised storage systems like **IPFS** or **cloud services**, the **control** over the security of that data diminishes. While DLT ensures tamper-proof storage for on-ledger data, off-ledger data may be susceptible to tampering, deletion, or unauthorised access, depending on the security measures employed by the external storage system.
    - **Tampering**: Off-ledger data, if not properly secured, can be altered by malicious actors. Since this data is stored outside the blockchain, it doesn’t automatically benefit from the immutability or consensus mechanisms of the DLT itself, leaving it vulnerable to manipulation.
    - **Data Integrity Risks**: If off-ledger data is tampered with, there must be mechanisms in place to ensure its integrity and verify that the data linked to the blockchain remains consistent and untampered. The inability to validate the integrity of off-ledger data poses a significant risk in applications where trust is critical, such as supply chain tracking or legal contracts.

Security for Off-Ledger Data

Given the risks associated with **on-ledger data**, particularly around public exposure, many enterprises store sensitive information **off-ledger**. However, off-ledger storage introduces its own set of security risks and considerations.

1. **Risks of Storing Data Off-Ledger**:

- **Loss of Control**: Once data is stored off-ledger, especially in external databases or decentralised storage systems like **IPFS** or **cloud services**, the **control** over the security of that data diminishes. While DLT ensures tamper-proof storage for on-ledger data, off-ledger data may be susceptible to tampering, deletion, or unauthorised access, depending on the security measures employed by the external storage system.
- **Tampering**: Off-ledger data, if not properly secured, can be altered by malicious actors. Since this data is stored outside the blockchain, it doesn't automatically benefit from the immutability or consensus mechanisms of the DLT itself, leaving it vulnerable to manipulation.
- **Data Integrity Risks**: If off-ledger data is tampered with, there must be mechanisms in place to ensure its integrity and verify that the data linked to the blockchain remains consistent and untampered. In applications where trust is critical (such as legally bidning smart contracts and supply chain tracking) the inability to validate the integrtity of the off-ledger data poses a significant risk.

Linking On-Ledger and Off-Ledger Data

The challenge of ensuring data integrity and security while balancing transparency and privacy is compounded when **on-ledger** and **off-ledger data** need to be linked. The solution often lies in using **cryptographic proofs** and **hashing** techniques to securely reference off-ledger data without storing the actual data on the ledger.

1. **Methods for Securely Linking Off-Ledger Data to On-Ledger Transactions**:
   - **Cryptographic Hashes**: A common method for linking off-ledger data to on-ledger transactions is the use of cryptographic **hash functions**. Instead of storing the actual data on the ledger, a **hash** (or "fingerprint") of the off-ledger data is stored. This allows any participant to verify the integrity of the off-ledger data without exposing the data itself.
   - **Merkle Trees**: **Merkle Trees** are used to efficiently organise and verify large datasets. By hashing each piece of off-ledger data and combining the hashes in a hierarchical structure, a **Merkle root** is generated and stored on the ledger. Any change in the off-ledger data would result in a different Merkle root, making tampering immediately detectable. This approach allows for scalable and efficient verification of off-ledger data while maintaining the security guarantees of on-ledger transactions.
2. **Ensuring that Off-Ledger Data Remains Verifiable and Tamper-Proof**:
   - **Tamper-Proofing**: Cryptographic techniques such as **digital signatures** and **hash functions** ensure that any modification of the off-ledger data is easily detectable. For example, if a document stored off-ledger is tampered with, the hash stored on the ledger will no longer match, alerting participants to the change.
   - **Proof of Existence**: The concept of **proof of existence** allows participants to prove that off-ledger data existed at a specific point in time without revealing the data itself. This is particularly useful in legal applications where sensitive data needs to be verified without being exposed. Cryptographic techniques like **zero-knowledge proofs (ZKPs)** can further enhance privacy while maintaining the verifiability of off-ledger data.
3. **Performance Considerations**:
   - Linking **on-ledger** and **off-ledger data** introduces potential performance challenges, particularly in large-scale DLT networks. The computational overhead involved in generating and verifying cryptographic proofs (e.g., hashes, Merkle Trees) can slow down the network, especially when handling high volumes of transactions. Solutions like **Layer 2** scaling techniques or **sidechains** can help alleviate these performance bottlenecks while maintaining the security and integrity of the on-ledger and off-ledger linkages.

Addressing Security Challenges for On-Ledger and Off-Ledger Data

The balance between **on-ledger** and **off-ledger data storage** presents complex security challenges for organisations adopting **distributed ledger technologies (DLTs)**. While **on-ledger data** benefits from cryptographic integrity and immutability, it is exposed to privacy risks, particularly in **public DLTs**, where data is visible to anyone participating in the network. To mitigate these risks, many organisations choose to store sensitive or large datasets **off-ledger**, but this approach introduces its own set of challenges, including risks of tampering, unauthorised access, and data integrity.

Security Vulnerabilities in Off-Ledger Storage: IPFS and Storj

1. **IPFS Security Vulnerabilities**:
   - **InterPlanetary File System (IPFS)** is a decentralised protocol used to store off-ledger data. However, while IPFS offers distributed storage, it suffers from known **security vulnerabilities**. One issue is the **lack of encryption by default**, meaning that data stored in

IPFS is not automatically protected. Anyone who has the **content identifier (CID)**, a unique cryptographic hash representing the file, can access the stored data.

- Additionally, **content persistence** in IPFS depends on node operators continuing to host the content. If no node voluntarily chooses to host the file, it can become unavailable, leading to potential **data loss**. Furthermore, there have been instances where **IPFS gateways** (which allow access to IPFS files via the HTTP protocol) were targeted in **DDoS attacks**, making data temporarily inaccessible.

2. **Storj Data Issues**:
   - **Storj**, another decentralised cloud storage platform, addresses some of IPFS's limitations by encrypting files before they are distributed across the network. However, **Storj** still faces issues related to **data durability** and **retrieval**. Because data is stored in pieces across multiple nodes, there have been reports of **data retrieval failures** due to network connectivity issues or insufficient replication of the file pieces. Moreover, while Storj does use encryption, the reliance on distributed nodes introduces potential attack vectors where the availability of data can be affected if the network is not sufficiently decentralised or robust.
   - Both IPFS and Storj illustrate that while **off-ledger data storage** offers privacy advantages, it also introduces risks related to data availability, persistence, and tamper-proofing, which organisations must carefully consider when choosing off-chain storage solutions.

Historical Examples of Blockchain Data Leaks

Despite the security guarantees offered by blockchain technology, there have been incidents where data stored on-chain or linked to the chain has been compromised:

1. **Mt. Gox Hack (2014)**:
   - One of the most infamous early **blockchain data breaches** occurred with **Mt. Gox**, a Bitcoin exchange. In 2014, hackers exploited weaknesses in Mt. Gox's systems, stealing over 850,000 bitcoins. While this breach was primarily a result of insecure operational controls rather than the blockchain itself, it revealed the risks of **on-chain data linkage** to centralised services that hold **private keys**. The stolen bitcoins were traceable on the blockchain, but due to poor off-chain security, they couldn't be recovered, demonstrating how vulnerable off-ledger data (such as private keys) can impact blockchain security.
2. **Bitcoin Deanonymisation Attack (2013)**:
   - In 2013, researchers showed that through analysing patterns of transaction flows and metadatait it is possible to **de-anonymise Bitcoin transactions** . By correlating public on-chain transaction data with off-chain data, attackers could trace Bitcoin addresses to real-world identities. This immediatelly highlighted the limitations of **pseudonymity** on public blockchains and the potential risks of data exposure, even if the on-chain data itself is cryptographically secure.

Organisations Improving Security for On-Ledger Data

Given the privacy risks associated with **on-ledger data**—particularly in **public DLTs**—several organisations have focused on improving the security of on-chain data to ensure that even if it exists in the public domain, it cannot be revealed.

1. **Zcash (Zero-Knowledge Proofs)**:
   - **Zcash** is a cryptocurrency that utilises **zero-knowledge proofs (ZKPs)** to enhance privacy on-chain. ZKPs allow one party to prove to another that a transaction is valid without revealing any details about the transaction itself (e.g., amounts, sender, and receiver). The innovation is that the data stored on the ledger remais private while still allowing it to be verifiable and auditable.
   - Zcash's use of **zk-SNARKs** (Zero-Knowledge Succinct Non-Interactive Arguments of Knowledge) is one of the most advanced implementations of privacy-preserving technology on a public ledger. This technology is particularly promising for businesses looking to store sensitive transaction data on-chain without risking exposure.
2. **Monero (Ring Signatures and Confidential Transactions)**:
   - **Monero** is another privacy-focused cryptocurrency that uses **ring signatures** and **confidential transactions** to obscure the details of on-chain data. Ring signatures allow a group of possible signers to approve a transaction without revealing which member of the group was the actual signer. Meanwhile, confidential transactions hide the amounts being transacted on the blockchain, adding an extra layer of privacy.
   - Monero's approach is regarded as one of the most robust in the blockchain industry. It ensures **on-ledger privacy** and is often being closely studied by many companies seeking solutions for buidling **private on-chain data storage**.
3. **Ethereum Layer 2 Solutions (zk-Rollups)**:

- **zk-Rollups**, a **Layer 2 scaling solution** on **Ethereum**, bundle multiple transactions off-chain and submit a single cryptographic proof to the main chain. This makes the Ethereum network able to maintain the integrity of the transactions without exposing individual transaction details. By processing transactions off-chain and only storing minimal proofs on-chain, zk-Rollups provide both privacy and scalability. To many startups and companies in the are this is for sure an attractive solution (especially for the ones handling large transaction volumes).

The Big Question is 'Can Personal Data Be Publicly Stored and Secured'?

The main concern for businesses revolves around whether **personal data** can be stored **in the public domain** (on-chain) in a manner that ensures it remains encrypted and secure, with no risk of exposure.

1. **Fully Homomorphic Encryption (FHE)**:
   - **Fully Homomorphic Encryption (FHE)** allows computations to be performed on encrypted data without needing to decrypt it. This means that data can remain encrypted even while being processed on a blockchain. While FHE holds enormous potential for securely handling **on-chain data**, it is still in its early stages of development, and current implementations are computationally expensive.
2. **Zero-Knowledge Proofs (ZKPs)**:
   - As seen with **Zcash** and **zk-Rollups**, ZKPs allow sensitive data to remain hidden while still being provable on-chain. This robust technology is increasingly being explored for many use cases beyond cryptocurrency - including **identity verification** and **supply chain management**. They require personal data can be encrypted and proven without being exposed.
3. **Confidential Computing (TEEs)**:
   - **Trusted Execution Environments (TEEs)** provide a hardware-based solution to ensure that sensitive data remains secure while being processed. TEEs isolate sensitive data from the rest of the system, preventing access by even the operating system or the cloud provider. TEEs, combined with DLTs, allow for private data to be processed and verified securely without being exposed to the public domain.

Security Innovations Are Advancing, But Trade-Offs Remain

The challenge of balancing transparency and privacy in **on-ledger** and **off-ledger data** storage remains complex. While **off-ledger storage** offers advantages in protecting sensitive data, it introduces its own set of risks, including potential tampering and data loss, as seen with platforms like **IPFS** and **Storj**. At the same time, keeping sensitive data **on-ledger** poses the risk of public exposure, especially in public DLTs, where deanonymisation attacks have proven that on-chain data is not always as private as it seems.

As we saw projects like **Zcash**, **Monero**, and Ethereum's **zk-Rollups** are pioneering solutions that improve privacy for on-chain data and offering ways to store data publicly while ensuring it remains secure. However, the question of whether **personal data** can be stored **in the public domain** while remaining encrypted and entirely unrevealable is still evolving. **Fully homomorphic encryption** and **zero-knowledge proofs** often offer promising avenues, but these technologies are still in development, and businesses must weigh the trade-offs between when the technology will be fully operation, its security, performance, and of course privacy.

Interoperability in Blockchain and Distributed Ledger Technologies

The interaction between on-ledger and off-ledger data, as previously described, is key in interoperability protocols. Ensuring seamless data exchange between these types requires unified standards for security and verification.

For years standards are designed around interoperability and it remains one of the most pressing issues in **blockchain** and **distributed ledger technology (DLT)** ecosystems. This is mainly trueto organisations I spoke with who grapple with how to store and manage data **on-ledger** and **off-ledger**. The ability of different blockchain systems—and blockchain systems with legacy IT infrastructures—to communicate seamlessly is essential. It will help achieving the necessary widespread adoption and efficiency that DLT promises. However, interoperability is a very ling path - it is still underdeveloped, with many challenges that need to be addressed through the adoption of common standards and protocols. While it remains a part of the domain of ISO/TC 307/WG 7, the work in this group is way from finished.

Importance of Interoperability in Blockchain

Interoperability the 'oly grail' of the standardisation. It ensures that **blockchain networks**, **applications**, and **systems** can exchange data, validate transactions, and work together without friction. This capability is vital for several reasons:

1. **Multi-Chain Ecosystems**: As organisations deploy different DLT platforms (e.g., Ethereum, Hyperledger Fabric, Corda), **cross-chain communication** is becoming essential. Without interoperability, the value of these systems is limited to the specific network in which they operate. **Interoperability** allows data and assets to flow freely between blockchains, promoting **scalability**, **flexibility**, and **collaboration** across ecosystems.
2. **Blockchain and Legacy Systems**: Many enterprises still rely on **legacy systems** for data management, such as **ERP systems** or **cloud-based databases**. Seamless integration between **blockchains** and these legacy infrastructures is essential for smooth data flow and efficient operations. The ability to move data between a blockchain and existing systems, particularly when storing **off-ledger data** or accessing on-chain records, is key to unlocking blockchain's potential.
3. **Cross-Industry Adoption**: Sectors like **finance**, **supply chain**, and **healthcare** depend heavily on industry-specific standards. Interoperability is critical to ensuring that blockchain technologies can comply with regulations and integrate with industry-specific protocols (e.g., ISO standards, GDPR compliance frameworks).

Interoperability Challenges in On-Ledger/Off-Ledger Data

The interoperability of **on-ledger** and **off-ledger data** introduces additional complexity, as different **blockchain systems** handle data storage and classification differently, and their security protocols often vary.

1. **Data Classification**: Different blockchains have varied approaches to data storage. For example, **Corda** uses a **state-based system** for transactions, while **Ethereum** relies on **smart contracts** and a more generalised ledger. These differences make it challenging to ensure seamless data exchange, especially when part of the data is stored **off-ledger**. Off-ledger storage solutions, such as **IPFS** or **cloud storage**, have their own security and integrity protocols, making it difficult to maintain a unified security and verification framework across systems.
2. **Security Protocols**: Security mechanisms (e.g., **encryption standards**, **access controls**) differ across blockchain platforms. When data is stored on-ledger in one system and off-ledger in another, ensuring that the data remains secure and tamper-proof across both environments can be difficult. A lack of standardised **security protocols** means that data could be vulnerable when transferred between systems, especially if cryptographic methods differ or are not interoperable.
3. **Cross-Chain Integration**: Moving data between **public blockchains** (e.g., Ethereum) and **private DLTs** (e.g., Hyperledger) presents significant challenges. Each chain has its own method of managing on-ledger and off-ledger data, making cross-chain verification difficult. This is particularly relevant in use cases like **supply chain management**, where data may need to be verified across different blockchain networks.

Standards and Protocols for Interoperability

Several existing **standards** and **protocols** are working towards achieving interoperability between systems, but the landscape is still fragmented, and no universal framework has been widely adopted.

6. **API Standards**: Application Programming Interfaces (APIs) are one of the most common methods of enabling interoperability between blockchain systems and legacy IT infrastructures. APIs allow data to be exchanged and processed by multiple systems, but they lack the standardisation necessary to ensure consistent handling of on-ledger and off-ledger data across different platforms.
7. **Cross-Chain Interoperability Frameworks**:
    - **Polkadot** and **Cosmos** are two prominent blockchain projects focused on building interoperability frameworks for cross-chain communication. They enable the transfer of assets and data between different blockchains, but the complexity of managing off-ledger data across chains remains largely unresolved.
    - **Wanchain** also works towards blockchain interoperability, allowing for cross-chain smart contract execution. However, the **integration of off-ledger data** still poses several challenges mainly in maintaining verifiability and security across different networks.
8. **ISO TC 307 WG7**: The **ISO TC 307** working group, particularly **WG7** on interoperability, is actively working on establishing standards to address these challenges. While progress is being made, no **technical standard** has yet been finalised that can be widely implemented across industries and blockchain platforms. The lack of a completed comprehensive standard leaves organisations struggling to develop

their own solutions or work with many proprietary technologies, which can limit the scalability and adoption of the distributed ledger technology as a whole.

9. **Suggestions for New Standards**:
    - With this report I see a form of a pressing need for **unified security protocols** for managing **on-ledger** and **off-ledger data** across systems. These protocols should standardise the use of **cryptographic proofs**, **hashing algorithms**, and **encryption methods** across different chains and off-chain storage systems to ensure data integrity and security.
    - **Cross-chain verification standards** are essential to ensure that data stored off-ledger in one system can be verified by another system using on-ledger cryptographic methods. This could be facilitated through the use of **universal hashing algorithms** or **multi-chain verification protocols** that allow systems to validate the integrity of off-ledger data even when operating across different blockchains.

Case Studies: Interoperability Solutions in Blockchain Ecosystems

1. Polkadot: A Multi-Chain Ecosystem with Cross-Chain Interoperability

Polkadot is a leading example of a **multi-chain blockchain** that focuses on **interoperability** by allowing multiple blockchains (called **parachains**) to operate within a shared security structure. Polkadot's **Relay Chain** acts as the central hub that manages cross-chain transactions and **data exchange** between parachains.

- **On-Ledger/Off-Ledger Data Handling**: Polkadot's parachains can store **on-ledger data** and interact with **off-ledger data** through bridges to external blockchains or off-chain storage systems. For instance, parachains can use **IPFS** or similar off-chain data storage mechanisms to handle large datasets, while still verifying and referencing these datasets on the Polkadot ledger via **hashing** or **cryptographic proofs**.
- **Interoperability**: Polkadot's core advantage is its ability to support **cross-chain interoperability**. Using **XCMP (Cross-Chain Message Passing)**, Polkadot enables parachains to communicate with each other and external blockchains. This interoperability is crucial for applications where different chains manage distinct types of data (e.g., financial transactions, supply chain records, and identity information), all of which may include both **on-ledger** and **off-ledger data**.
- **Challenges**: While Polkadot has made significant progress in **cross-chain interoperability**, managing off-ledger data across multiple parachains presents challenges. Polkadot ensures data integrity through cryptographic hashes stored on the ledger, but the complexity arises when different parachains have varied data security protocols, making unified data integrity verification across parachains difficult to achieve.
- **Takeaway**: Polkadot provides an advanced framework for interoperability, but the technical challenge of linking **on-ledger** and **off-ledger data** securely and efficiently across different parachains remains an area requiring further development.

2. Cosmos: The Internet of Blockchains

Cosmos aims to solve blockchain fragmentation by creating a network of blockchains that can communicate and transact with one another using the **Inter-Blockchain Communication (IBC) protocol**. Cosmos focuses on enabling **independent blockchains** (or "zones") to interact securely and seamlessly.

- **On-Ledger/Off-Ledger Data Handling**: In the Cosmos ecosystem, each **zone** can have its own method of handling **on-ledger** and **off-ledger data**. The Cosmos **Tendermint Core** consensus engine ensures the integrity of on-ledger transactions, while off-ledger data can be stored externally, with cryptographic hashes linked to the blockchain. This allows Cosmos to handle large datasets off-chain while verifying the data's existence and integrity on-chain.
- **Interoperability via IBC**: The **IBC protocol** is Cosmos's solution for interoperability, facilitating the secure transfer of tokens and data between different blockchains. This protocol is particularly valuable for systems that manage both **on-ledger** and **off-ledger data**, as it allows for the secure movement of assets or information between blockchains with different data handling approaches. For instance, a Cosmos zone managing **on-chain financial transactions** can interact with a supply chain zone that stores **off-ledger inventory data**.
- **Challenges**: The challenge lies in ensuring that **off-ledger data** linked to **IBC transactions** remains verifiable across different blockchains. While IBC enables cross-chain communication, it does not directly address the issue of ensuring **data integrity** for off-ledger data across different chains, leaving this responsibility to each individual zone.
- **Takeaway**: Cosmos provides a scalable solution for blockchain interoperability through its IBC protocol, but the management of **off-ledger data** remains decentralised and dependent on the specific zones, leading to potential security and performance inconsistencies.

3. Hyperledger Fabric: Enterprise-Grade Interoperability with Legacy Systems

Hyperledger Fabric, as an enterprise-focused blockchain framework, is often used in **supply chain**, **financial services**, and **healthcare** industries. Hyperledger Fabric's modular architecture allows it to integrate with legacy systems and other blockchains while maintaining secure data storage.

- **On-Ledger/Off-Ledger Data Handling**: Hyperledger Fabric supports **private data collections** for sensitive or off-ledger data. This allows organisations to store data off-chain while still referencing it on-chain through **cryptographic hashes**. Fabric ensures that data stored off-chain remains accessible to authorised parties only, thanks to robust **access control policies** embedded in its architecture.
- **Interoperability with Legacy Systems**: Hyperledger Fabric excels in integrating **blockchain data** with **legacy enterprise systems**. Through its modular framework, Fabric allows data to flow between **on-chain ledgers** and **off-chain systems** such as **ERP** and **CRM** platforms. For example, a manufacturer may store product lifecycle information in its ERP system while referencing key transaction details on the Hyperledger Fabric blockchain. This allows for **auditability** without the need to store large datasets directly on-chain.
- **Challenges**: While Fabric's modular design facilitates integration with external systems, the lack of a universal **cross-chain interoperability protocol** makes it difficult to interact with other blockchain networks. Fabric can integrate with legacy systems easily, but cross-chain communication—especially for off-ledger data across different DLTs—remains limited without external frameworks.
- **Takeaway**: Hyperledger Fabric offers powerful interoperability with **legacy enterprise systems** and can manage off-ledger data securely through private collections. However, cross-chain interoperability with other blockchain platforms remains underdeveloped, limiting its utility in multi-chain ecosystems.

4. Chainlink: Cross-Chain Interoperability and Data Oracles

Chainlink is a decentralised oracle network that connects blockchains with external data sources, allowing smart contracts to interact with **off-chain data** in a secure and reliable manner. Chainlink provides **interoperability solutions** betweenvarious blockchains and also between blockchains and **real-world data sources**.

- **On-Ledger/Off-Ledger Data Handling**: Chainlink operates as a **data oracle**, feeding **off-ledger data** (such as financial market prices, weather data, and IoT sensor information) into **on-ledger smart contracts**. Chainlink uses **decentralised oracles** to gather and verify external data before making it available on the blockchain. This off-chain data is linked to on-chain contracts via **cryptographic proofs**, ensuring data integrity and preventing tampering.
- **Cross-Chain Interoperability**: Chainlink has extended its functionality to enable **cross-chain data exchanges** through **Chainlink Cross-Chain Interoperability Protocol (CCIP)**. This protocol is designed that different blockchains can securely exchange data and assets (the aim is to solve one of the core challenges of **cross-chain interoperability)**. By ensuring that both **on-ledger** and **off-ledger data** can be securely exchanged between blockchains, Chainlink enables complex multi-chain ecosystems where data needs to move seamlessly.
- **Challenges**: The main challenge for Chainlink is ensuring that **off-chain data sources** are trustworthy. While Chainlink's decentralised oracles reduce reliance on a single point of failure, the quality and accuracy of external data remain a risk. Additionally, integrating Chainlink with multiple blockchains requires ongoing adjustments to ensure compatibility with different network architectures.
- **Takeaway**: Chainlink offers a strong framework for **on-ledger and off-ledger data interoperability**, particularly in enabling smart contracts to interact with real-world data. However, the challenge of ensuring the integrity of off-chain data sources and maintaining compatibility across diverse blockchains remains a key focus area.

5. Quant Network: Blockchain Interoperability Pioneer

Quant Network is at the forefront of blockchain interoperability with its Overledger platform, designed to seamlessly connect different blockchain networks and legacy systems. Overledger enables enterprises, including those in industries like finance, healthcare, and supply chain, to exchange and synchronise on-ledger and off-ledger data across multiple blockchain ecosystems without the need for intermediaries.

- **On-Ledger/Off-Ledger Data Handling**: Overledger allows for the storage of critical transaction data on-ledger, while larger datasets or sensitive information can be maintained off-ledger in secure environments. The platform ensures that off-ledger data can still be referenced on-chain through cryptographic proofs, maintaining both privacy and verifiability.
- **Interoperability**: Overledger is unique in that it provides interoperability not only across blockchain networks (like Ethereum, Ripple, and Hyperledger) but also between blockchain systems and legacy enterprise infrastructure. This capability ensures that businesses can integrate blockchain without disrupting their existing IT systems.

- **Challenges**: Quant Network's biggest challenge lies in achieving widespread adoption of its interoperability protocols across diverse industries. While Overledger is highly flexible, convincing legacy systems and stakeholders to trust and integrate the technology remains a hurdle. Additionally, the regulatory environment for cross-chain transactions is still evolving, which could impact the platform's scalability.
- **Takeaway**: Quant Network exemplifies the potential of blockchain interoperability beyond just supply chains, offering a solution that can bridge both blockchain ecosystems and traditional enterprise systems. Its Overledger platform ensures secure and efficient data exchange, though widespread adoption and regulatory concerns present challenges.

Note: I only have a limited research time and with the so many blockchain projects out there it is impossible for me to cover every interoperability effort. Sincere apologies to the ones that I failed to include in the report, so feel free to email me at petko.karamothcev@industria.tech and I will make sure to do whatever possible to include the missing pieces of the puzzle in the later edition of the report.

Interoperability as the Cornerstone for Blockchain's Maturity

Interoperability is critical for the widespread adoption and scalability of **blockchain ecosystems**. Yet, the current landscape reveals that progress has been slower than anticipated. Initiatives such as **ISO TC 307 WG7**, and platforms like **Polkadot** and **Cosmos**, have made strides in developing frameworks, but there is still a significant gap in achieving fully functional interoperability protocols that can effectively manage **on-ledger** and **off-ledger data**. This gap has tangible consequences: enterprises face barriers in scaling blockchain solutions across multiple systems, integrating with legacy infrastructures, and ensuring data security when linking on-ledger and off-ledger storage.

Concrete Steps Toward Blockchain Interoperability

To overcome these challenges, it is not enough to recognise the need for interoperability. Specific steps must be taken to lay a realistic path forward for blockchain systems to achieve seamless cross-chain and cross-system communication:

1. **Adoption of Modular Standards for Data Management**:
   - The immediate priority for interoperability is the creation and widespread adoption of **modular standards** that address the differing methods of **on-ledger** and **off-ledger data management**. This includes standardising how data is hashed, encrypted, and validated across different blockchains and off-chain storage solutions.
   - A clear example of progress in this direction is the work being done by **ISO TC 307 WG7**. However, to be effective, the standards produced by this working group must transition from theoretical frameworks into widely adopted technical specifications. **ISO TC 307** should accelerate its efforts to produce actionable, universally applicable standards for handling **cross-chain** and **off-ledger data** interactions.
2. **Implementation of Cross-Chain Verification Protocols**:
   - A critical step is the development of **cross-chain verification protocols** that allow for the secure validation of **off-ledger data** between different blockchains. These protocols must enable chains with varying data structures (such as those using **smart contracts** or **state-based systems**) to communicate and validate data integrity, especially when data is stored off-chain. **Cosmos' IBC** protocol and **Chainlink's CCIP** offer early frameworks, but they need to be expanded with stronger data verification tools that can handle larger, more complex datasets stored off-ledger.
   - Projects like **Polkadot's XCMP** have already started to address **message-passing** between chains, but the inclusion of off-ledger data into these protocols is still in its infancy. Ensuring that **cross-chain** messaging can validate both **on-ledger** and **off-ledger data** is paramount for enterprise adoption.
3. **Establishing Unified Security Standards**:
   - A major concern for enterprises using blockchain solutions is the consistency of **security protocols** across multiple chains and data environments. Current blockchain ecosystems operate with varied security protocols, particularly when managing **off-ledger data**. Without a unified approach, data flowing between blockchains and **legacy systems** remains vulnerable to security breaches and compliance issues.
   - **Unified encryption standards**, **hashing algorithms**, and **access control policies** must be developed, tested, and enforced across all blockchain platforms to ensure data security. Blockchain networks and **decentralised storage providers** (such as **IPFS** or **Storj**) must adhere to these standards to ensure that data stored off-chain can be securely linked to on-ledger data without risking exposure or tampering.

4. **Incentivising Industry Collaboration**:
    - Blockchain interoperability will not succeed without collaboration between **enterprises**, **standards organisations**, and **technology providers**. Blockchain platforms, particularly **public networks** like **Ethereum** and **Bitcoin**, must collaborate with **private DLTs** like **Hyperledger Fabric** and **Corda** to create interoperable solutions. Public-private partnerships can also accelerate the adoption of standards that benefit both sectors, enabling smoother data flow between **on-chain** and **off-chain** environments (although we all know how difficult the bridge between innovative forward thinkers and standard writers like me is).
    - Governments and regulatory bodies (can you imagine, there are 49 government bodies and a few international organisations working on ISO/TC 307 - Blockchain and distributed ledger technologies) should incentivise cross-industry collaboration, offering funding or regulatory sandboxes for projects that focus on developing **interoperability standards**. This would help bring blockchain systems into mainstream enterprise use, promoting innovations in **supply chain**, **healthcare**, **finance**, and **IoT**.
5. **Investing in Interoperability Research and Development**:
    - Research into **Layer 2 solutions** (such as **zk-Rollups**) and **Layer 3 protocols** is critical for scaling interoperable blockchain systems. While Layer 2 focuses on improving scalability and performance, Layer 3 must focus on creating **interoperability bridges** that ensure both **on-ledger** and **off-ledger data** can be seamlessly integrated across chains.
    - Industry stakeholders must prioritise R&D investment in **multi-chain architecture**. The future of blockchain interoperability depends on the ability to execute cross-chain smart contracts and manage off-ledger data while ensuring that performance does not degrade as complexity increases.

A Path Forward for Interoperability

Interoperability is no longer just a buzz word, an item in a presentation deck or a requested feature—it is a **strategic imperative** for blockchain to fulfil its promise of transforming industries. The blockchain community, including **standards bodies**, **platform developers**, and **industry stakeholders**, must work very hard and commit to delivering practical, deployable interoperability solutions. This is a wake up call for them that these solutions must be be anchored in well-defined **technical standards**, robust **cross-chain verification protocols**. Unified **security frameworks** that ensure data integrity and privacy across all systems are also necessary.

The path forward is clear: interoperability requires **1) Modular Standards**, **2) Cross-chain communication tools**, **3) Security Unification**, and, most importantly, **4) Collaboration** between industries and technology providers. This should be not a very distant goal, but one that can be achieved with focused efforts in the coming years. Enterprises that invest in these developments now will be positioned to leverage blockchain's full potential as a secure, scalable, and interoperable technology.

Legal Enforceability

On-chain and off-chain records are indeed related to the **legality** and **enforceability** of smart contracts, but they play different roles in the broader context of a legal system. This is a favourite topic of mine and here's how they interact and influence the legality of a smart contract:

1. On-Chain Records:

These are the parts of the smart contract and associated data stored directly on the blockchain. They are immutable and publicly verifiable, offering a clear record of the contract's terms, executions, and associated events. The **on-chain** component provides key legal advantages:

- **Proof of Transaction**: Blockchain's immutability and transparency provide a clear, auditable trail of actions such as payments, property transfers, or contract modifications.
- **Self-Execution**: Smart contracts can automatically enforce pre-defined terms without human intervention (e.g., payment schedules or penalties for default).
- **Digital Signatures**: Parties can sign the contract digitally, and the blockchain records that signature, adding an extra layer of proof.

Legal Relevance:

- **Proof of Agreement**: The on-chain record can be presented as proof of a contract's existence and the parties' consent to its terms. However, traditional courts may require more than just on-chain data to deem a contract enforceable, depending on the jurisdiction.
- **Enforcement of Terms**: Since smart contracts can execute predefined rules automatically, the on-chain data provides enforceability within the blockchain ecosystem. However, for broader legal recognition (e.g., in case of disputes), human interpretation may still be

required for ambiguous or subjective clauses.

- **Regulatory Compliance**: On-chain data can be critical in adhering to regulations, like anti-money laundering (AML) and know-your-customer (KYC) compliance, as it ensures that specific actions are properly recorded.

2. Off-Chain Records:

These refer to data and records that are stored off the blockchain, either for technical reasons (e.g., scalability issues) or due to sensitive or private information not suited for public blockchains. Off-chain records often include:

- **Private Legal Agreements**: Some legal aspects of a contract, like certain personal data, negotiation terms, or confidential business details, may remain off-chain.
- **Supporting Documents**: External documents (such as mortgage applications, employment verifications, or property deeds) are often stored off-chain but referenced within the smart contract.
- **Oracles**: External sources of truth (e.g., interest rates, legal rulings, or economic indicators) that inform the smart contract of off-chain events are essential but exist outside the blockchain ecosystem.

Legal Relevance:

- **Complete Legal Context**: Off-chain records provide the full legal context that courts may require when interpreting a contract. For example, while on-chain data may show a financial transaction, off-chain data might include explanations or reasons for contract modifications or disputes.
- **Human Interpretation**: Traditional legal systems are built around contracts that require human interpretation for ambiguous terms, subjective judgments, or unforeseen circumstances. Off-chain records often play a significant role in legal disputes, as they provide flexibility that on-chain records lack.
- **Regulatory Concerns**: Off-chain records are often critical for ensuring that all regulatory requirements are met, as not all compliance can be enforced purely through on-chain mechanisms. For example, GDPR compliance for personal data might require off-chain storage and handling.

The Relationship Between On-Chain and Off-Chain Records:

The interaction between on-chain and off-chain records determines how a smart contract can be legally enforced in the real world:

- **Hybrid Contracts**: Many legal smart contracts will adopt a **hybrid model**, where core terms (payment schedules, interest rates, penalties) are on-chain, while more complex, discretionary, or personal clauses remain off-chain. This approach allows for automation and transparency while maintaining flexibility for human interpretation.
- **Legal Enforceability**: Courts and legal systems may require access to both on-chain and off-chain data to assess the full scope of the contract. An **on-chain smart contract** by itself may not be considered legally binding in every jurisdiction, especially if critical details are off-chain. Having off-chain records can help ensure that the contract aligns with local laws and regulations.
- **Dispute Resolution**: In the event of disputes, the **on-chain record** provides a clear and immutable trail of what happened, but **off-chain documents** are often necessary to resolve ambiguities, subjective claims, or external conditions that the on-chain contract wasn't programmed to handle.

Proposed Work Item

New Framework for Classifying On-Ledger/Off-Ledger Data

In order to address the current challenges surrounding **on-ledger** and **off-ledger data management** within blockchain ecosystems, we propose the development of a comprehensive **classification framework**. This framework will provide clear guidelines for classifying and deciding when data should reside **on-ledger** (directly on the blockchain) and when it should be stored **off-ledger** (in external or decentralised storage systems). The aim is to standardise the criteria for data classification, while we seek to ensure scalability, security, and regulatory compliance across different blockchain platforms and protocols.

1. **Development of a Clear Classification System**:

- The proposed classification framework should categorise data based on factors such as **size**, **privacy requirements**, **transaction frequency**, **regulatory needs**, and **security implications**. For example, data that is essential for **public transparency** and **auditability** (such as financial transactions or smart contract logic) would be classified as **on-ledger**, while large files, sensitive personal data, or dynamic information that changes frequently (such as IoT sensor data) would be classified as **off-ledger**.
- **Data Sensitivity Levels**: The framework should introduce levels of data sensitivity, defining what qualifies as sensitive, confidential, or public data, to help organisations make informed decisions about where to store specific types of information.
- **Impact of Regulations**: The framework will incorporate global regulatory requirements, such as **GDPR**, **HIPAA**, and **eIDAS**, to ensure that compliance is considered when classifying and storing data. For instance, personal data requiring deletion under GDPR could automatically be designated as **off-ledger** data, while ensuring the integrity of the data using cryptographic techniques on-chain.

2. **Steps for Standardising On-Ledger/Off-Ledger Data Storage**:
   - **Criteria Development**: The framework will establish a set of universal criteria to determine whether data should be stored **on-ledger** or **off-ledger**. These criteria would include:
     - **Data Size**: Due to storage limitations and costs on blockchain networks, larger datasets should be stored off-ledger, with only critical metadata or cryptographic hashes stored on the ledger.
     - **Transaction Frequency**: High-frequency data, such as IoT sensor readings or real-time business metrics, should remain off-ledger due to the dynamic nature of the data, with periodic snapshots or proofs being stored on-ledger.
     - **Privacy and Compliance**: Personal data, intellectual property, and proprietary business data will typically remain off-ledger to comply with privacy laws and protect sensitive information.
   - **Integration of Cryptographic Proofs**: The framework will advocate for the use of **cryptographic proofs** and **Merkle Trees** to link off-ledger data with on-ledger transactions. This will ensure that off-ledger data can be verified and audited without compromising the necessary levels of security or privacy.
3. **Standardisation Through ISO TC 307**:
   - The proposed classification framework will align with existing efforts under **ISO TC 307**, particularly working group **WG7**, which is focused on blockchain interoperability. By integrating this new framework into ISO TC 307 standards, we can ensure that it becomes a widely adopted practice across blockchain ecosystems.
   - This work item will also include recommendations for technical standards on **API integration** and **cross-chain data validation** that allow seamless interaction between on-ledger and off-ledger data in different blockchain systems, enhancing the interoperability between various platforms.

Some Recommendations for Future Work

While the report you are currently reading addresses the (almost) immediate need for classifying **on-ledger** and **off-ledger data**, there are several areas for further research and development to ensure continuous improvement and scalability.

1. **Further Research Areas**:
   - **Privacy-Preserving Techniques for On-Ledger Data**: Future work should explore how modern techniques, such as **fully homomorphic encryption** and **zero-knowledge proofs**, can enable personal and sensitive data to be stored on-ledger without exposing it to public eyes. This will clearly expand the range of data that can be stored on-chain while maintaining privacy and security.
   - **Scaling Solutions for Off-Ledger Data**: Research into improving **off-ledger storage** solutions, such as **decentralised storage networks** (e.g., IPFS, Storj) and **cloud-based systems**, to enhance their **reliability**, **security**, and **performance** in large-scale implementations. Even as blockchain networks continue to grow, off-ledger data should remains accessible, secure, and tamper-proof
2. **Potential Collaborations and Partnerships**:
   - **Industry Collaboration**: Engage with industry groups such as **Hyperledger**, **Ethereum Enterprise Alliance**, and **R3 Corda** to refine and test the classification framework in real-world blockchain applications. My expectations for these groups to provide feedback on how the framework operates in practice and help improve it for different use cases, including supply chain, healthcare, and finance.
   - **Academic Partnerships**: Collaboration with academic institutions to further the development of innovative techniques for linking **on-ledger** and **off-ledger data** securely. Universities with strong cryptography and distributed systems research programs can help drive

advancements in this area, particularly regarding new privacy-preserving techniques.

- **Regulatory Engagement**: Work with global regulatory bodies (e.g., **EU Data Protection Board**, **Financial Action Task Force**) to ensure that the classification system and standards align with **data protection laws** and **financial compliance regulations**. This engagement will ensure that the framework not only meets technical requirements but also operates within legal guidelines.

Expert Insights on On-Ledger and Off-Ledger Data

In addition to the research and analysis conducted, we facilitated a series of discussions and surveys with **23 recognised experts** from various sectors. I needed their contribution to focus on real, practical experience, their challenges, and give me insights and opportunities regarding the management of distributed ledger data. These insights should provide a broader industry perspective and could validate the necessity for clear standards in this area.

Key Themes from Expert Contributions:

1. **Data Classification and Management:** The consensus among the experts was that effective data classification between on-ledger and off-ledger storage is essential for the scalable adoption of blockchain technology.
   - **Transaction Data and Public Verifiability**: 78% of experts agreed that crucial data for public verifiability, such as transaction records and smart contract execution, should be stored on-ledger to leverage immutability and transparency. However, they warned that excessive data storage on-ledger could burden networks, citing Ethereum's gas fees as an example of high-cost implications for on-ledger storage.
   - **Off-Ledger for Efficiency**: Experts recommended that large datasets, such as multimedia files or IoT sensor data, and personal data should be stored off-ledger, using blockchain to store only cryptographic hashes or metadata for verification purposes.
2. **Security and Privacy Considerations:** Security and privacy were central concerns, particularly with regards to off-ledger data.
   - **Cryptographic Methods**: 65% of participants highlighted the importance of using cryptographic methods, such as digital signatures and zero-knowledge proofs (ZKPs), to ensure the integrity of off-ledger data while maintaining the immutability of the on-ledger records.
   - **GDPR Compliance**: Experts from the legal sector, making up 30% of the group, were particularly concerned with regulatory compliance, specifically GDPR. They strongly advised that personal data and sensitive business information remain off-ledger to align with data protection and privacy laws.
3. **Cost-Efficiency and Scalability:**
   - **On-Ledger Cost Limitations**: Around 70% of the experts raised concerns over the cost of storing large data sets on-chain. They noted that public blockchains such as Ethereum could quickly become impractical for organisations that generate vast amounts of data, pushing organisations to find hybrid solutions.
   - **Off-Ledger Solutions**: Decentralised storage networks, such as IPFS and Filecoin, were frequently mentioned as viable off-ledger options to reduce costs while maintaining data verifiability. This approach allows for cost-efficient scalability, where only essential data resides on-chain.
4. **Interoperability as a Major Challenge:**
   - **Lack of Cross-Chain Standards**: A resounding 74% of experts expressed frustration over the lack of standardisation and interoperability across blockchain platforms. They advocated for the development of universal standards to enable secure data exchange between different chains, particularly when managing off-ledger data.
   - **Enterprise-Grade Solutions**: Blockchain architects in the expert group highlighted that while enterprise-grade platforms (e.g., Hyperledger Fabric, R3 Corda) are advancing, interoperability remains a barrier for systems that need to exchange on-ledger and off-ledger data across industries.
5. **Regulatory and Legal Implications:**
   - **Compliance-Driven Data Decisions**: 60% of the experts, particularly those from regulated industries (e.g., healthcare, finance), stressed the importance of blockchain systems being adaptable to regulatory changes. The right to be forgotten (a GDPR requirement) was noted as problematic for immutable on-ledger data, suggesting that hybrid on/off-ledger strategies should be pursued to balance compliance with blockchain's inherent immutability.

Expert Contributions Summary

To strengthen the report with practical, real-world insights, the following table summarises the inputs of 23 blockchain experts, categorised by industry, expertise, and their specific contributions to the discussion of on-ledger and off-ledger data management.

To comply with **GDPR** regulations, the names of the participating experts have been *anonymized*. However, their participation and the information provided have been thoroughly verified. All experts have given their explicit consent, stating:

> *"I consent that my participation is solely for the purposes of this initiative, and my responses will be used accordingly."*

The table below summarises the key insights gathered from the expert panel, categorized by industry and area of expertise.

Expert	Industry	Area of Expertise	Key Insight/Comment
Expert 1	Finance	Blockchain in Finance	Advocates for on-ledger use for critical financial records to ensure trust
Expert 2	Healthcare	Data Security & Privacy	Recommends off-ledger for PII to comply with GDPR and minimise security risks
Expert 3	Legal	Regulatory Compliance	Stresses the importance of privacy laws, advocates off-ledger for personal data
Expert 4	Supply Chain	Blockchain in Supply Chain	Suggests hybrid on/off-ledger solutions for tracking and cost optimisation
Expert 5	Cybersecurity	Cryptographic Techniques	Emphasises the use of digital signatures and hashing for off-ledger security
Expert 6	Enterprise Technology	Cloud Computing & Blockchain	Highlights the need for off-ledger storage for scalability and data retrieval
Expert 7	Government	Public Blockchain Policy	Supports on-ledger for transparency but suggests off-ledger for sensitive data
Expert 8	Real Estate	Smart Contracts & Ownership Rights	Advocates for on-ledger in ownership and property rights records
Expert 9	Energy	Decentralised Energy Markets	Stresses scalability via off-ledger storage of energy consumption data

Expert 10	Retail	Blockchain for Retail	Highlights cost-efficiency of off-ledger for large, frequently changing datasets
Expert 11	Academia	Blockchain Research & Development	Recommends research into Layer 2 scalability solutions to optimise on-chain storage
Expert 12	IoT	IoT Integration with Blockchain	Advises off-ledger storage for IoT sensor data due to its volume and frequency
Expert 13	Legal	Intellectual Property Management	Off-ledger for securing IP, on-ledger for verifiable proof of ownership
Expert 14	Banking	Digital Currencies	Supports on-ledger for transactional integrity, warns of high costs for storing large data sets
Expert 15	Healthcare	Patient Data & Compliance	Recommends off-ledger storage for healthcare records, on-ledger for metadata
Expert 16	IT Infrastructure	Blockchain System Design	Suggests hybrid solutions for large-scale data, emphasises off-ledger for scalability
Expert 17	Manufacturing	Blockchain in Supply Chain	Strong support for off-ledger in tracking large volumes of raw material data
Expert 18	Insurance	Claims Processing & Blockchain	Suggests on-ledger for fraud prevention but off-ledger for storing large claims files
Expert 19	Finance	DeFi & Asset Tokenisation	Recommends on-ledger for tokenised assets, off-ledger for supplementary documents
Expert 20	Telecom	Blockchain in Telecommunications	Supports off-ledger for large, frequently updated telecom logs
Expert 21	Entertainment	NFTs & Blockchain	Emphasises the need for off-ledger storage for

			media files linked to NFTs
Expert 22	Logistics	Fleet Management & Tracking	Advocates hybrid on/off-ledger approach for tracking goods and compliance data
Expert 23	Supply Chain	Blockchain in Supply Chain	Highlights off-ledger to ensure cost-effective and scalable visibility

Note: I would like to express sincere thanks for all who participated in the interview and promise to shorten and simplify my questions in the next edition of a similar interviews. For the sake of transparency, one interviewed reported that the level of technical depth is too big for their current level of knowledge.

Synthesis of Expert Contributions and Key Takeaways

The insights gathered from this diverse pool of experts underline the current complexity of managing on-ledger and off-ledger data. While there is a general agreement on the importance of on-ledger immutability for specific types of data, **experts strongly advocate for off-ledger solutions** where scalability, cost, and privacy are key concerns.

The insights gathered from our experts have shaped a gentle recommendation for undertaking a balanced approach to data management in blockchain and distributed ledger. In this approach, cryptographic methods sould securely connect off-ledger data to on-ledger transactions. The result will be a promoted scalability while ensuring compliance with regulations. I believe that as blockchain technology evolves, these expert contributions could become a key in setting up additional standardisation work that will encourage the foreseen wider adoption.

Conclusion

Effectively managing on-ledger and off-ledger data is critical for ensuring that blockchain systems remain scalable, secure, and compliant with regulations, while also retaining legal integrity.

Key Takeaways from Expert Input

The discussions with industry experts was necessary and it highlighted the urgent need to continue developing comprehensive standards. These standards must address data classification, security, and interoperability across on-ledger and off-ledger storage in blockchain and distributed ledger technologies. The consensus among the experts is clear: as blockchain systems scale, the decision of where data resides —whether on-chain or off-chain—directly impacts privacy, security, scalability, and adherence to regulations like GDPR and HIPAA.

On-Ledger Data: The consensus is that on-ledger data should be used primarily for immutable, verifiable records, such as transaction logs, smart contract executions, and proof of ownership. This data must be easily auditable and trusted by all network participants. However, storing large or sensitive information directly on the blockchain comes with risks such as privacy concerns, higher costs, and limits to scalability.

Off-Ledger Data: Off-ledger data, while providing flexibility for larger datasets or sensitive information, introduces concerns around integrity and security, particularly when the off-ledger data is linked back to the blockchain. Many experts highlighted the need for robust cryptographic techniques to ensure that off-ledger data remains tamper-proof and accessible while alleviating the strain on blockchain resources.

The expert panel also confirmed that **interoperability** between blockchain platforms and legacy systems remains one of the biggest challenges to widespread blockchain adoption. The fragmentation between systems limits blockchain's potential, especially in enterprise environments where seamless data exchange is critical. The role of ISO TC 307 and working groups like WG7 on interoperability is seen as pivotal in addressing these issues.

Next Steps Informed by Expert Insights

Finalising the Data Classification Framework: The experts agree that a structured framework is necessary to help businesses make informed decisions about which data should reside on-ledger versus off-ledger. As technology experts, our mission is to maintain the balance between the blockchain usefulness in the shape of immutability and transparency and the necessities of the enterprise for privacy, scalability, legal valididity and cost-efficiency. The input from this expert survey reinforces the need of aligning this framework with the ongoing ISO TC 307 work to drive global adoption.

Developing Cross-Chain and Cross-System Interoperability Protocols: The results from the discussions with the expert highlighted the fragmented nature of blockchain ecosystems. We can now clearly see a real need for new interoperability protocols that can handle both on-ledger and off-ledger data across different systems. Experts emphasized the importance of cross-chain verification mechanisms that can ensure the integrity of off-ledger data when referencing on-chain.

Enhancing Unified Security Standards: A consistent theme from the expert feedback is the need for unified security protocols that govern both on-ledger and off-ledger data. Encryption, cryptographic integrity, and secure access control mechanisms are essential to protect sensitive data and ensure its verifiability. Collaboration between blockchain networks and decentralised storage platforms, such as IPFS or cloud-based systems, is necessary to fortify data security.

Promoting Collaboration Between Industry and Standards Bodies: The expert panel highlighted that industry collaboration is key to advancing standardisation efforts. Public-private partnerships can, could and should be encouraged, especially to test new standards based on real-world blockchain deployments. Governments and regulatory bodies can also incentivise these collaborations through sandboxes and funding initiatives.

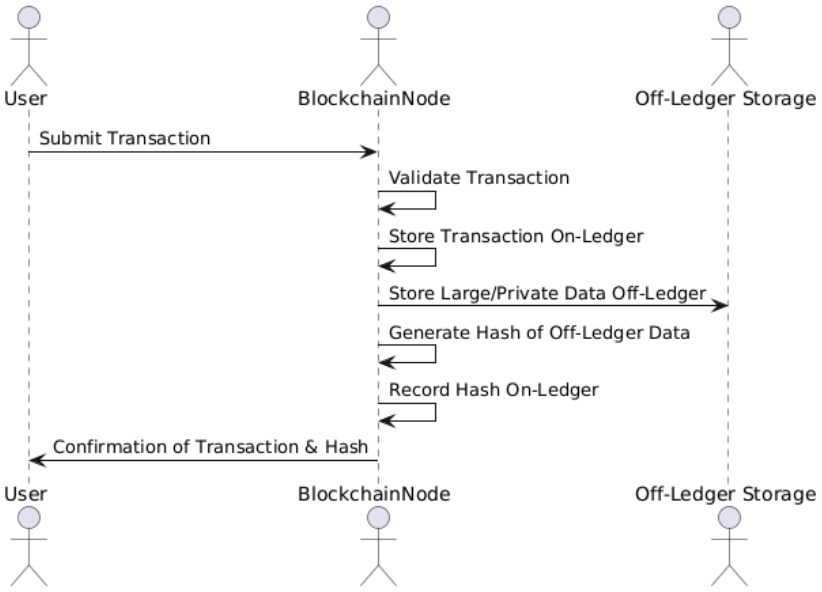
Investing in Privacy-Preserving Technologies and Off-Ledger Scalability: The experts pointed out that research into privacy-preserving technologies like zero-knowledge proofs (ZKPs) and fully homomorphic encryption (FHE) should be accelerated to enable secure on-chain storage of sensitive data. Furhter development of decentralised off-ledger storage solutions is needed to address performance, availability, and security challenges.

Conclusion

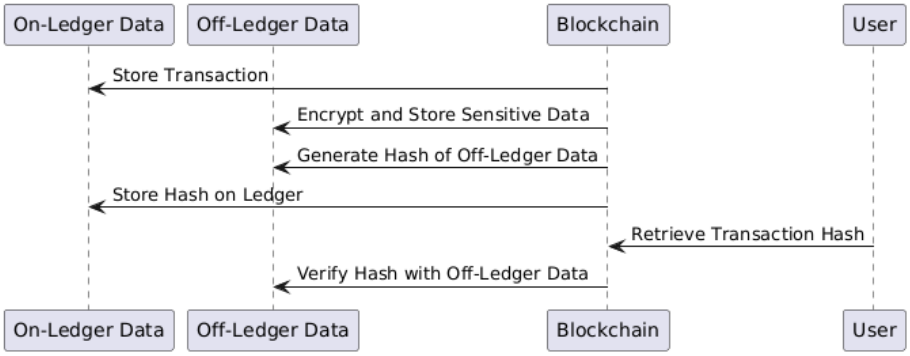
I am happy that the expert insights reaffirmed my views of **the critical role that 1) Proper Data Classification, 2) Security, 3) Interoperability,** and **4) Legal Enforceability** play in the future of blockchain and distributed ledger technology. The findings from the expert survey validate the proposed recommendations and demonstrate a clear path forward for improving blockchain scalability, security, and enterprise adoption through standardisation and collaboration. We need to continue the research and te development of unified standards. Through them the blockchain ecosystem will be better equiped and will handle the galloping demands of the glogal industry.

Appendices

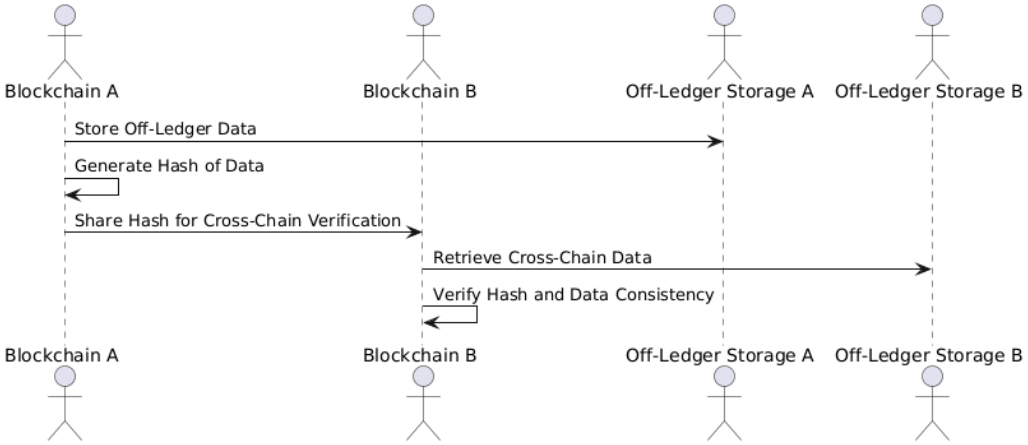
Technical Diagrams



On-Ledger vs. Off-Ledger Data Flow



Security Mechanisms for On-Ledger and Off-Ledger Data



Interoperability Protocols for On-Ledger and Off-Ledger Systems

Glossary of Terms

- **Blockchain**: A decentralised ledger of all transactions across a network of nodes. Blockchain ensures immutability, transparency, and security by using cryptographic techniques.
- **On-Ledger Data**: Data that is stored directly on the blockchain. This data is immutable and benefits from the security and transparency provided by the decentralised network.
- **Off-Ledger Data**: Data that is stored externally (in decentralised or centralised systems) but linked to the blockchain via cryptographic hashes or proofs. Off-ledger data is used when handling large datasets, private information, or when high transaction frequency is required.
- **Cryptographic Hash**: A fixed-size string of characters generated from input data of arbitrary size, used to ensure data integrity. In blockchain, hashes are stored on-chain to represent off-ledger data.
- **Merkle Tree**: A hierarchical data structure used in blockchains to efficiently and securely verify data integrity. Each leaf node is a hash of data, and every non-leaf node is a hash of its children.
- **Zero-Knowledge Proof (ZKP)**: A cryptographic method by which one party can prove to another party that a statement is true, without revealing the actual information that supports the statement.
- **Fully Homomorphic Encryption (FHE)**: A form of encryption allowing computations to be performed on ciphertext, generating an encrypted result that, when decrypted, matches the result of operations performed on the plaintext.
- **Interoperability**: The ability of different blockchain networks, or blockchains and legacy systems, to interact and exchange data in a seamless, secure manner.
- **Layer 2 Scaling**: Off-chain solutions (like Lightning Network for Bitcoin or Optimistic Rollups for Ethereum) that process transactions off the main blockchain (Layer 1) to reduce congestion and enhance scalability.
- **Distributed Ledger Technology (DLT)**: A decentralised database managed by multiple participants, typically across various locations. Blockchain is a subset of DLT.
- **Cryptographic Signature**: A mathematical scheme for verifying the authenticity of digital messages or documents. A valid signature ensures that the message comes from the claimed sender and hasn’t been altered.
- **ISO/TC 307**: An International Organization for Standardization (ISO) committee focused on the development of standards for blockchain and distributed ledger technologies.
- **Decentralised Storage (IPFS/Storj)**: Systems that store data across multiple nodes in a decentralised manner, enabling redundancy and reduced reliance on centralised servers.