Blockchain and SDMX: Immutable data sharing for trust

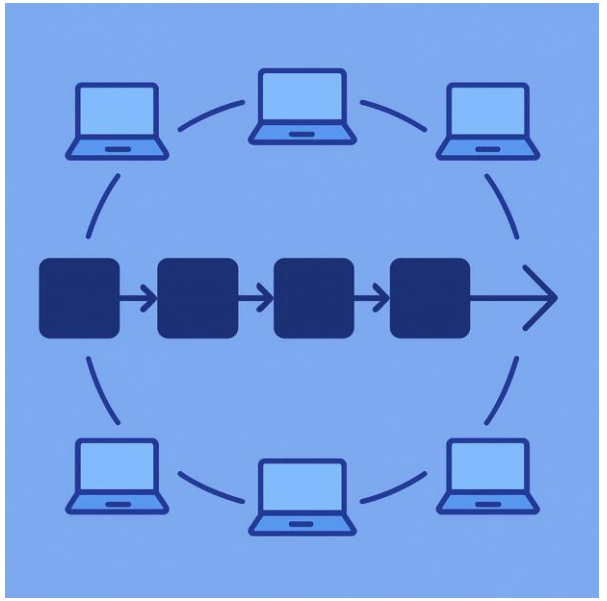
SDMX Global Conference

Edward Lambe, Glenn Tice, Rafael Schmidt, Christian Schmieder (BIS) and Mario Rusev (d-fine)

Agenda

- Blockchain for Statistical Data Exchange
- Why Blockchain & SDMX Solves Real Problems
- Why now?
- Two Use Cases with Clear Business Value
- Implementation proposal
- Pilot implementation
- Discussion and feedback

Blockchain for Statistical Data Exchange



A blockchain is a decentralized, tamper-resistant digital ledger that chronologically records transactions in cryptographically linked "blocks" across a network of participants, ensuring security and transparency without relying on a central authority

Key Features:

- **Decentralization:** Eliminates single points of failure and control by distributing data
- **Transparency & Auditability:** Complete logged transaction history to participants, fostering trust
- **Security:** Data immutability and cryptographic hashing
- **Consensus Mechanisms:** Protocols like Proof-of-Work and Proof-of-Stake empower the network to validate new blocks collaboratively, ensuring unanimous agreement on a state without intermediaries
- **Common Use Cases:** Cryptocurrencies and Digital Assets, Smart Contracts, Supply Chain Management, Secure Record-Keeping

Benefits:

- Increased Trust, Enhanced Security, Traceability and Auditability, Greater Efficiency and Cost Reduction

Why Blockchain & SDMX Solves Real Problems

- **Builds Global Trust:**
  - Provides a ***transparent, cryptographically secure and verifiable record*** of official statistics, enhancing credibility globally
  - Strengthens international confidence through verifiably transparent data dissemination practices
- **Ensures Data Integrity:**
  - ***Guarantees immutability and reliability of statistics*** via cryptographic hashing and consensus mechanisms, establishing the foundation for evidence-based policymaking
  - Significantly mitigates risks of data manipulation and maintains high-quality standards
- **Promotes Interoperability via SDMX** *(Verification Recommended)*:
  - Blockchain can inherently incorporate SDMX standards, facilitating effective, robust and standardized international data exchange.
  - Enhances **cross-system compatibility and data comparability** across global statistics
  - Allows **instant** data validation compared to legacy public key infrastructure

Why now?

- **Need for Evidence-Based Policy Making:**
  - Policymakers depend critically on accurate, timely, and verifiably credible data to formulate robust and impactful decisions.
  - Blockchain establishes an immutable chain of custody for data and metadata, guaranteeing authenticity and integrity and thereby creating a trusted foundation for mission-critical evidence-based policies.
- **Advances in AI and Machine Learning:**
  - Modern AI/ML models can process and analyze massive, complex datasets faster than ever before.
  - Blockchain serves as a critical enabler for trustworthy AI by providing a pipeline of high-fidelity, cryptographically assured data, significantly reducing the risks of bias, data poisoning and unreliable outputs in automated analysis.
- **Agent-Based Automation** *(Conceptual SDMX Integration Recommended)***:**
  - The statistical ecosystem appears to be moving towards autonomous, agent-based systems for end-to-end data processing, validation, and dissemination.
  - An SDMX-integrated blockchain architecture, could allow streamlining machine-executable workflows by natively embedding SDMX standards, ensuring efficiency, unwavering consistency, and verifiable reliability throughout the full data lifecycle.

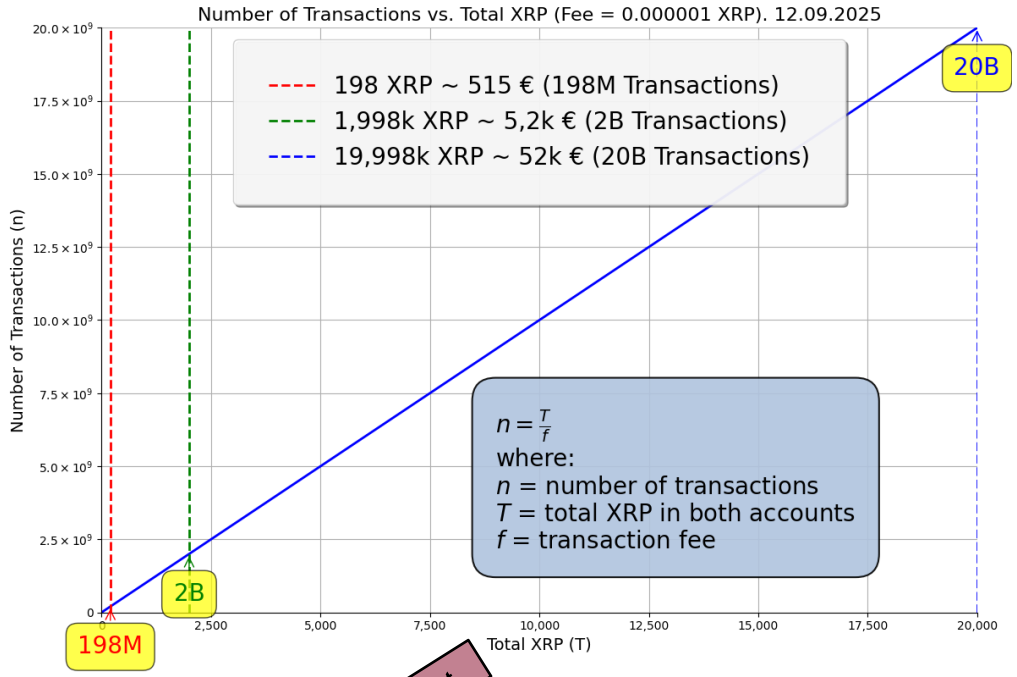
Two Use Cases with Clear Business Value

Two **sample** use cases to prevent the misuse of official statistics on unverified commercial or third-party platforms by cryptographically binding datasets to their origin:

- Use Case 1: **Immutable Version Authentication**
  - *Cryptographic Provenance*: Trace datasets unambiguously back to the official producer
  - *Immutable Audit Trail*: Record every version and modification with tamper-proof proofs
  - *Digital Signature Authentication*: Producers sign each dataset release, proving origin and integrity
  - *Real-Time Validation*: Enable end users to instantly verify the authenticity via lightweight cryptographic proofs
  - *Standards Alignment*: Support SDMX schemas and ensure GDPR compliance
- Use Case 2: **Redistribution Trust Framework**
  - *Portable Proofs*: Third-party platforms can bundle data with tamper-proof credentials
  - *Selective Disclosure*: Producers reveal only authorized fields via zero-knowledge proofs
  - *Blockchain Anchoring*: Immutable fingerprints of each dataset version are anchored to public blockchains
  - *Privacy-Preserving Revocation*: Manage credential revocations without revealing user identities

Associated costs – holdings and transaction capacity for our XRPL DevNet PoC

- Metainformation is written as a *memo*.
- Costs are kept at *minimum* by using two blockchain accounts, rotating sender/receiver if required, referencing only transaction fee (assuming full control of seeds).
- Possibilities to extend the PoC for smart contracts (*conditional transactions*) or NFTs (e.g. *ownership*), **inflation-linked products**, real-time markets, **transparent dashboards powered by immutable data**, perpetual futures markets, issuance of new types of digital assets etc.



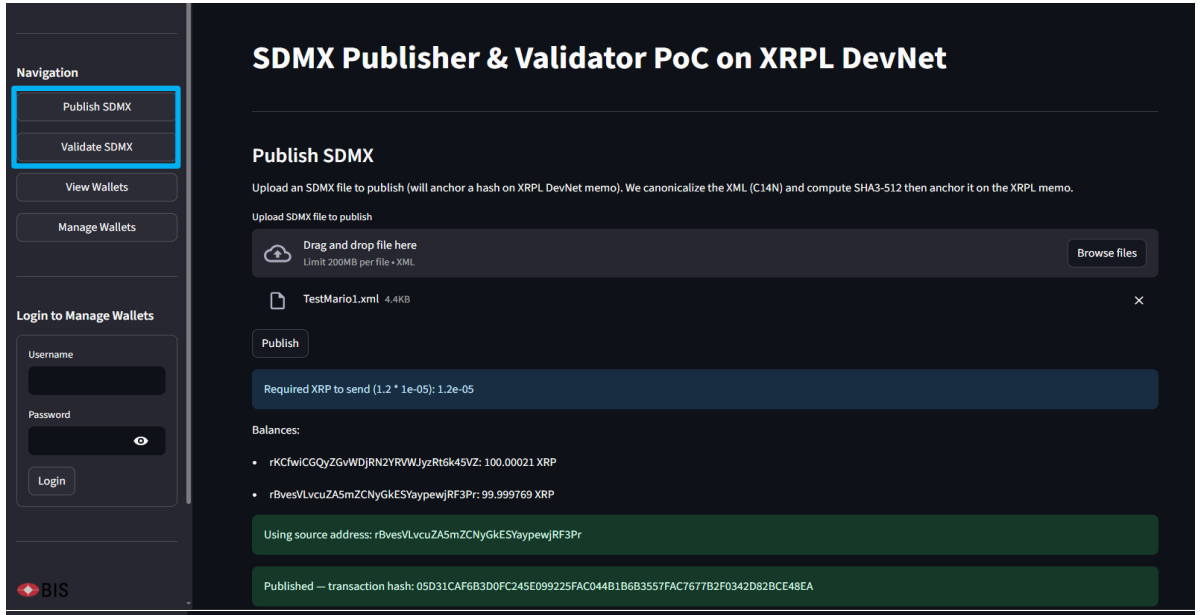
Personal Measurements

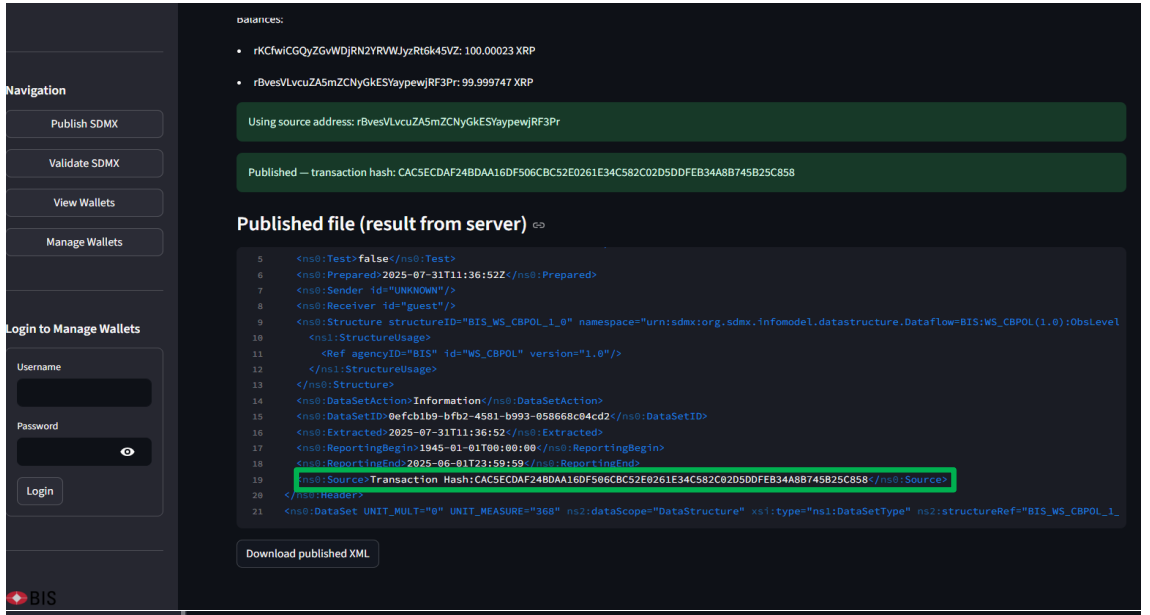
Feature	Our Implementation	Chainlink Approach	Ethereum Approach	Legacy Public Key Infrastructure
Verification Time	Fixed < 5s (ts validation), < 2s (query)	seconds to minutes (ts validation) < 2s (query)	several minutes (ts validation) < 4s (query)	Hours to days
Transaction fee	~$0.0000028	~$2 to $50	~$0.50	~$5.00 (Cert Authority cert)
Integration Complexity	Low (API-based)	Low (API-based)	Medium (custom integration required)	High (infrastructure)
Real-time Verification	Yes	Yes	Limited	No
SDMX Compliance	Full	Full (via metadata)	Variable	Variable

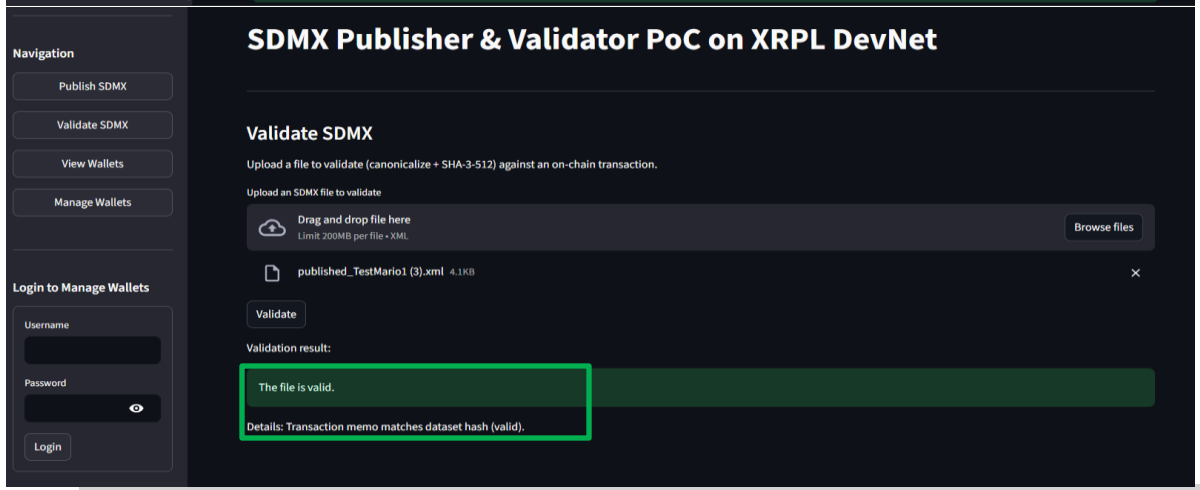
Local Tests + DevNet

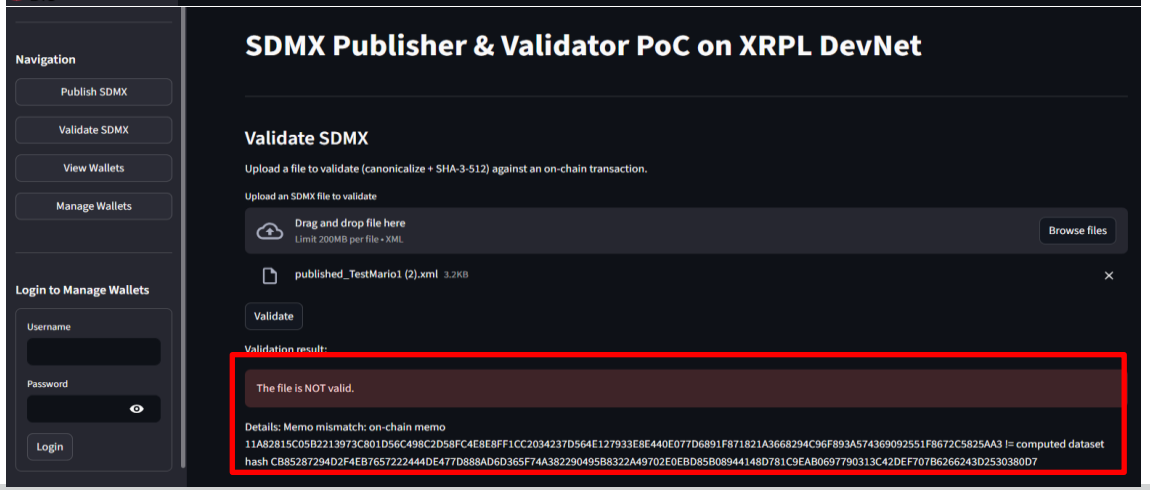
Feature	Timing
Compute SHA3-512	~ 4 ms
Canonicalize XML bytes (C14N)	~15 ms
Total End-To-End Publication Time and GUI Response	~5-7 s
Total Processing Time for a validation	~1 s

Pilot implementation demo









Discussion and feedback

- **Relevance and Priority**
  - Do you see a real need for authenticating SDMX data through blockchain in your institutional context?
  - Where do you rank this use case compared to other SDMX innovations?
- **Feasibility and Adoption**
  - What technical or organizational barriers would your institution face in adopting blockchain for data dissemination?
  - Would a decentralized validation model improve your trust in redistributed data?
- **Collaboration and Governance**
  - Which actors (e.g., NGOs, international organizations, private vendors) should maintain and govern the blockchain infrastructure or use a public blockchain?
  - Would embedding digital signatures or metadata hashes in SDMX messages be practical or too complex?
- **User Perspective and Trust**
  - Would authenticated data from third-party platforms increase public trust in official statistics?
  - How do you imagine users (e.g., journalists, researchers, businesses) would engage with blockchain-authenticated data?

Data trust survey

- Where do the threats lie?
- What technical options are preferred?
- What are the benefits of cryptographically verifiable data?
- Granularity of trust. Dataset, series, observation levels?
- Where does data trust figure in your organisational priorities?

Self-paced survey open until **October 9**

https://app.sli.do/event/jRrnP8ak2RNewFCkbEeouF