

FairOnChain: A FAIR Infrastructure for Blockchain Data Analytics

M Sas⁴, S Augusto², A Alwash², V Berisha², C Russo², M Thadthapong³, Ü Öztürk¹, S Casale-Brunet¹, C Lebrun², N Tovanich³, M Mattavelli¹, A Gaudinat², J Prat³, W J Knottenbelt⁴

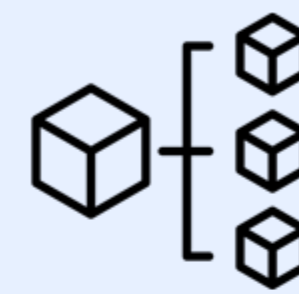
¹ École Polytechnique Fédérale de Lausanne
² Haute École de Gestion de Genève / HES-SO

³ CREST, École Polytechnique, Université Paris-Saclay
⁴ Imperial College London

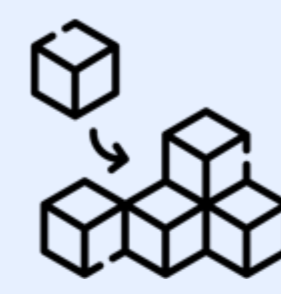
ABSTRACT

Blockchain data is inherently accessible, but difficult to analyze due to the lack of unified structure and standardisation. This project goal is to build an open infrastructure based on the FAIR principles to make blockchain data *Findable*, *Accessible*, *Interoperable* and *Reusable* for research purposes.

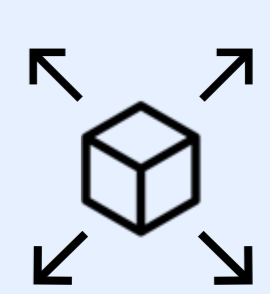
SYSTEM DESIGN & ARCHITECTURE



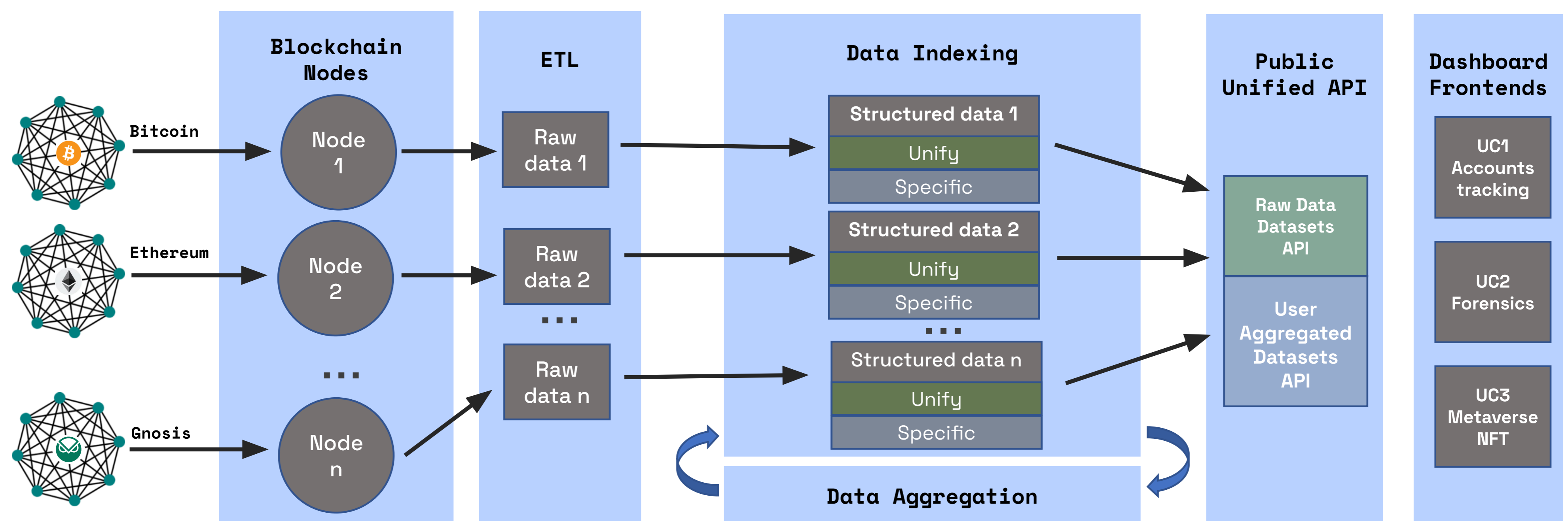
distributed



modular



extensible



METHODOLOGY

1. Deploy full blockchain nodes (Bitcoin, Ethereum etc)
2. Design modular ETL (Extract, Transform, Load) pipelines
3. Store collected data in unified, indexable formats
4. Develop advanced search tools for efficient querying
5. Share data in open formats (CSV, JSON, Parquet)
6. Expose standardized APIs for access & query
7. Distribute primary data via decentralised storage (IPFS)

FEATURES & INNOVATIONS

- **Metadata & provenance** using Dublin Core, schema.org & PROV-O
- **Structured datasets** following FAIR principles
- **Tiered access:** free for academics, licensed for commercial users
- **Governance** by a European academic consortium


DISSEMINATION & IMPACT

- Open access datasets and software tools
 - released under MIT, CC BY-NC-SA licenses
- Submission to standards bodies (ISO, Software Heritage)
- Awareness & community building via socials & website


CONCLUSION

FairOnChain provides the foundation for transparent and collaborative blockchain research in Europe and beyond, ensuring scientific reproducibility and open innovation.


KEY USE CASES

1. **Wallet & Smart Contract Activity Tracking** 

to analyse different groups of users in the blockchain & query accounts based on their behavioural profiles.

 - index transaction data related to EOAs (Externally-Owned Accounts) & CAs (smart Contract Accounts)
 - identify tokens implemented in smart contracts
 - extract DeFi protocol (decentralized finance) transactions
 - precompute account aggregate statistics (received, sent, balance)
2. **Cryptocurrency forensics** 

to allow regulators to see the relationships of money flow from one account to others.

 - provide fast access to transaction networks as graph data
 - develop a visual analytics system to help follow the money
 - identify who receives money after money laundering operations
3. **Metaverse & NFT Analytics** 

to study the economy & users' behaviour on trading & pricing strategies for NFT assets (Non-Fungible Tokens) across multiple blockchains.

 - integrate the data from both on-chain and off-chain
 - perform graph analysis on the transaction-level information
 - develop a visualization tool to help general audiences explore different tokens and who owns them across blockchains

RESULTS & PERFORMANCE METRICS

- Archival or full data in synch with **90-100%** of live nodes
 - for Ethereum, Bitcoin, Holesky, Gnosis, Lukso blockchains
- Query latency: **10-50 ms**
- **Six** public datasets discoverable via Google/OpenAIRE
- Launch of hackathons and workshops
- Peer-reviewed publications