TrialChain: A Blockchain Implementation for Decentralized Governance of Clinical Trial Data

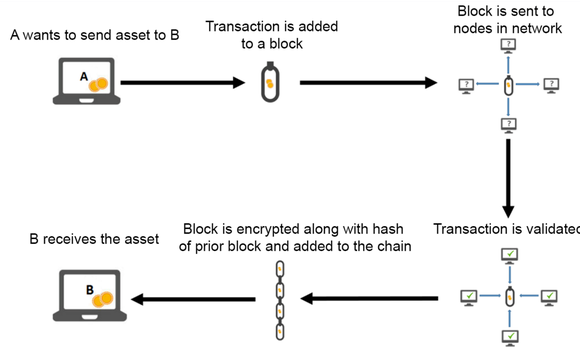
Wade L. Schulz, MD, PhD[1,2], H. Patrick Young, PhD[1], Hao Dai[3], Thomas JS Durant, MD[1,2], Lixin Jiang, MD, PhD[3], Harlan M. Krumholz, MD, SM[1,4]

[1]Yale New Haven Hospital, Center for Outcomes Research and Evaluation; 1 Church St #200, New Haven, CT 06510
[2]Yale University School of Medicine, Department of Laboratory Medicine; 55 Park Street PS345D, New Haven, CT 06511
[3]Fuwai Hospital, National Center for Cardiovascular Disease; Beijing, China
[4]Yale University School of Medicine, Department of Internal Medicine, 330 Cedar St, Boardman 110, New Haven CT, 06520

BACKGROUND

Data validation and data governance are key to any data science project, especially within regulated industries such as healthcare and finance. Within biomedical research, the prevalence of scientific fraud has been a recurring issue within both the academic and commercial sectors. While the improvement of study reproducibility and data transparency may require a multifaceted approach, the use of emerging cryptographic technologies may reduce the risk of fraudulent data practices and boost the confidence in conclusions made by the scientific community. The recent expansion of blockchain technology provides a novel approach that can be used to rapidly deploy cryptographically-secure data validation and audit trails with open source technology.

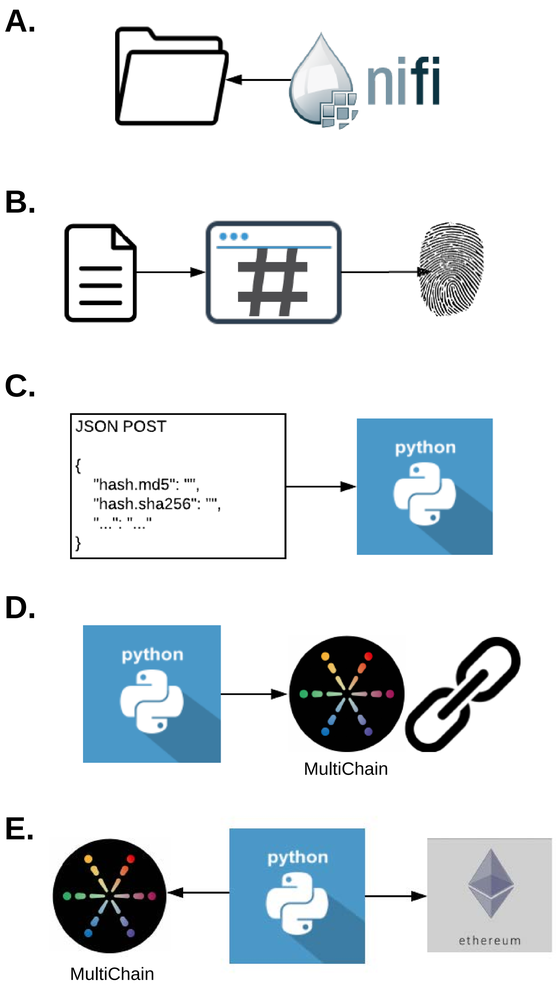
BLOCKCHAIN OVERVIEW



METHODS

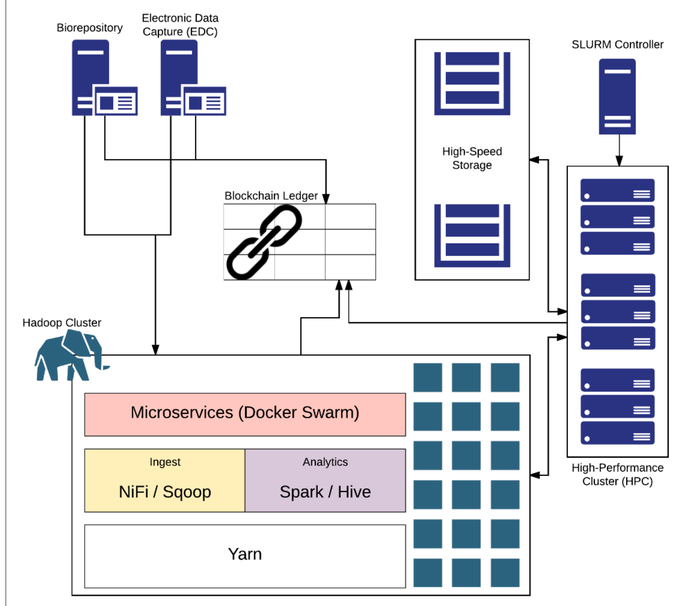
- MultiChain software (v1.0.2), based on the Bitcoin protocol, implemented as a secure, private blockchain
- Python (v3.5) web service interface created to receive custom data/audit messages for incorporation within the blockchain
- Geth (v1.7.3) deployed for integration with the public Ethereum network
- NiFi (v1.4.0) deployed with a custom workflow to hash files and provide metadata to Python web service endpoint
- Deployed all scripts and software within Docker containers to increase portability and security of the TrialChain infrastructure

IMPLEMENTATION RESULTS



TrialChain Workflow

- A NiFi workflow monitors source systems for new files/data assets (A) which are hashed (B) to create a cryptographic digital fingerprint of the file
- The hash, file metadata, and timestamp are then sent to a web service (C) which logs the data into a private blockchain (D)
- Periodically, a Python script obtains the current block hash from the private blockchain and submits it to the public Ethereum network (E) so file states and timestamps can be publicly audited



TrialChain Test Deployment

- Multi-million person trial to assess cardiovascular disease and risk factors throughout China
- Collection of structured and unstructured data, including surveys and whole exome sequencing
- Implemented data science platform based on Hadoop architecture and integrated high performance computing (HPC) infrastructure
- Active TrialChain deployment to create verifiable audit log for all data collected within the platform

CONCLUSIONS

- Private blockchains can be used to create cryptographically-secure data governance platforms when linked to public blockchains
- By linking only the block hashes between private/public systems, no private or confidential information ever leaves the private system
- Since block hashes are based on the content of all prior data and files, the state of any file can be publicly audited at any time
- While production deployments are needed to fully evaluate, blockchain technology may provide an efficient technical solution for data governance when there is a risk for data manipulation
- Potential to extend to additional use cases within clinical trials, such as validation of IRB approval, decentralized/patient-centered consent, and tracking changes in study protocols

ACKNOWLEDGEMENTS

We would like to acknowledge Suveen Angraal in the Center for Outcomes Research and Evaluation for discussions related to blockchain implementation and the adapted figure in the overview and Mingming in the National Center for Cardiovascular Disease for initial integration discussions to include next generation sequencing data within the platform.

REFERENCES

1. Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. 2008. (Accessed Januray 15, 2017, from: http://www.cryptovest.co.uk/resources/Bitcoin%20paper%20Original.pdf)
2. Angraal, S., et al. (2017). "Blockchain Technology: Applications in Healthcare." Circulation: Cardiovascular Quality and Outcomes 10(9).
3. Carlisle, B. (2014). "Proof of prespecified endpoints in medical research with the bitcoin blockchain." Retrieved 2018-02-04, from https://www.bgcarlisle.com/blog/2014/08/25/proof-of-prespecified-endpoints-in-medical-research-with-the-bitcoin-blockchain/.