LinkChains: Trusted Personal Linked Data

Conference or Workshop Item

LinkChains: Trusted Personal Linked Data

Allan Third[1 [0000−0002−0386−1936]] and John Domingue[1 [0000−0001−8439−0293]⋆]

Knowledge Media Institute, The Open University, Milton Keynes MK7 6AA, UK
{allan.third,john.domingue}@open.ac.uk

Abstract. Centralisation of data is increasingly widely considered to be an important societal problem. The ability to manage personal data, especially, in accordance with good principles for decentralised data systems is vital in response. We present our vision for the trusted decentralised personal data Web in the form of the *LinkChains* architecture.

Keywords: Blockchain · Semantic Web · Linked Data · Solid.

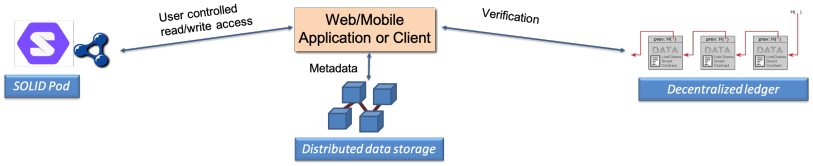
Centralisation on the Web has been a topic of increasing concern recently, with a corresponding interest in decentralisation. In particular, there are concerns about personal data in large volumes aggregated in the control of a single entity who is not the subject of the data. Our vision is of a world where individuals are in control of their own personal semantic data, on decentralised platforms following the FAIR TRADE principles [5] – in particular, where the content of such data can be verified and trusted by third parties wherever needed.

Centralisation of data can have global effects. The potential influence of Cambridge Analytica on recent major political events was reported as a major scandal [1], for example – made possible by the quantities of personal data held and controlled by a single entity, Facebook. These concerns are not restricted solely to the digital domain either. The loss of centralised records and distrust of non-centrally-held records led to 63 British citizens being deported, and up to 57 000 more being at risk of losing their UK rights and residency [2], [6].

Issues such as these have prompted initiatives such as Solid [4], which aims to offer personal Linked Data "pods", allowing users to hold and control their own data. From another direction, the concept of blockchains [7] arose as a *distributed ledger*; a shared append-only record of immutable securely timestamped entries.

In [5], we presented "FAIR TRADE", an extension of the FAIR principles [8] for data management to cover approaches to decentralised data. **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable are extended with **TR**usted, **A**utonomous, **D**istributed, and d**E**centralised. We present LinkChains as a FAIR TRADE platform for personal Linked Data using blockchains to provide a trust backend. LinkChains offers user privacy and control, and granular data verification. A cryptographic hash of metadata relating to personal data is stored in an immutable blockchain record, signed by a trusted party. Such a hash serves as a pointer to the metadata location on decentralised

⋆ Supported by QualiChain H2020 ICT 822404.



file storage platforms such as IPFS [3] capable of storing larger volumes of data more easily and cheaply than most blockchain platforms. The immutability of the blockchain record, and the nature of hashing, ensures the timestamp, immutability, and immunity from faking of the IPFS data. An individual's Solid pod contains their actual data, as Linked Data where possible, with pointers both to the blockchain and IPFS records. Relevant hashes of data and metadata in each location can be collectively used to verify the publication and untampered status of each.

LinkChains is thus a fully decentralised platform which supports individual hosting and control of their personal data while maintaining the ability for external parties to verify its contents where needed. Further extension to LinkChains involves the use of Merkle trees of subsets of the data, as fine-grained as terms or quads. The tree data ensures that queries can return verifiable proofs of data at the given granularity – we refer to this augmented query architecture as *MerQL*. The strength of the MerQL approach is with federation; proofs can be included with individual data items included in results aggregated from federated, sources.

[5] shows how LinkChains can support FAIR TRADE data management. In brief, the novel principles (TRADE) are supported since LinkChains data can be Trusted via the blockchain records and hashing, Autonomous due to individual control of personal data, and Distributed and dEcentralised due to the underlying Solid, distributed storage, and blockchain technology. LinkChains provides a research environment for exploring Linked Data and blockchains in general, and how notions of trust, identity, and decentralisation interact with each other, as well as enabling the implementation and study of the implications of the vision we opened with: of a world where individuals are in control of their own personal semantic data, on trusted FAIR TRADE decentralised platforms.

References

1. The Cambridge Analytica Files: the story so far, https://gu.com/p/8b2tc
2. Windrush generation, https://fullfact.org/immigration/windrush-generation/
3. Benet, J.: IPFS - Content-addressed, versioned, P2P file system. (2014), http://arxiv.org/abs/1407.3561
4. Berners-Lee, T.: Solid project website, https://solid.mit.edu/
5. Domingue, J., Third, A., Ramachandran, M.: The FAIR TRADE framework for assessing decentralised data solutions. In: Companion Proceedings of The 2019 World Wide Web Conference. pp. 866–882 (2019), http://oro.open.ac.uk/60149/
6. Grierson, J.: Windrush row: 63 peole could have been wrongly removed, https://gu.com/p/8j428
7. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system (2009)
8. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., et al.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific data **3** (2016)