Blockchain, web3 & Digital Humanities

Foteini Valeonti

Post-doctorate Fellow at the Sloane Lab (UCL Department of Information Studies)

TOWARDS A NATIONAL COLLECTION | UKRI Arts and Humanities Research Council | UCL | The British Museum | Natural History Museum | Technische Universität Darmstadt

The Sloane Lab: Looking back to build future shared collections

Collaboration between **UCL** and **TU Darmstadt**, the **British Museum**, the **Natural History Museum**

In partnership with: Archives and Records Association; Down County Museum; National Galleries of Scotland; Collecting the West; University of Oxford; British Library; Royal Botanic Garden Edinburgh; National Museums Scotland; and Historic Environment Scotland

Sloane Lab's goal: To reunite the collections of Sir Hans Sloane, i.e. the founding collections of the original British Museum (divided between the present-day British Museum, British Library & Natural History Museum)

Special study inspired by the needs of the Sloane Lab

Decentralising Digital Humanities: Exploring Blockchain Technology and 'web3' for the Sloane Lab and Towards a National Collection (TaNC)

Abstract

Purpose – Advancements in Internet technologies greatly influence digital humanities, yet research investigating web3 (i.e. the blockchain-based, decentralised web) within that domain remains limited. The purpose of this paper is to address that gap, presenting a state-of-the-art synthesis of web3-related technologies for digital humanities infrastructures and exploring associated risks and challenges.

Design/methodology/approach – Following a review of the literature, the authors scope out ways blockchain technology, peer-to-peer decentralised storage and other web3 technologies could support digital humanities infrastructures, especially in the context of digital cultural heritage. In this discussion, particular cognisance is given to the needs and aims of the UK's Arts and Humanities Research Council funded Towards a National Collection programme, which seeks to break down the barriers that exist between the UK's cultural heritage collections.

Findings – Web3 introduces novel tools and processes that could benefit digital humanities infrastructures, enabling decentralisation and facilitating open access data storage. Yet, significant barriers to adoption remain, such as the requirement for highly specialised technical expertise. Risks and challenges must also be considered prior to any use, including legal, ethical and technical safeguards.

Research limitations/implications – This study explores opportunities and risks of web3 for digital humanities, through the lens of digital cultural heritage infrastructures and their requirements,

(Under review)

What is a blockchain?

Definition: a distributed, immutable ledger that is maintained and verified among a network of peers

In simple terms: a commonly shared database that is accessible by everyone and cannot be compromised by anyone

What do blockchains do?

Blockchain technology enabled trust-less transactions between complete strangers *without the involvement of a trusted third-party*

Alice can send money to Bob without the oversight of a centralised authority

By removing the need for trusted intermediaries, blockchain technology made decentralisation possible

Emergence of ‘web3’

Whereas the first ever blockchain, i.e. Bitcoin, was focused on financial transactions, Ethereum in 2014 implemented Smart Contracts, i.e. autonomous software programmes operating on the blockchain

Fostered for the development of Decentralised Applications (DApps) and the advent of a new form of Internet that is based on decentralisation

Decentralising Digital Humanities

Although Internet is central in Digital Humanities, web3 has yet to be examined

We identify 4 challenges faced by digital humanities projects, web3 could potentially be of benefit, assessing associated risks

We also identify areas core to DH, where web3 is not expected to make an impact in the medium-term *(Digitisation & Data Analysis)*

Decentralising Digital Humanities

Main Findings:

Web3 is largely work in progress

For DH few solutions are immediately deployable and directly usable *Highly-specialised technical expertise is required*

However: significant potential and we should familiarise ourselves and be able to monitor the field

Web3: Potential areas of use for DH

1. **Data Storage** *(Cost & over-centralisation)*
2. **Persistent identification**
3. **Data permanence** *(Maintaining outputs post-grant)*
4. **Interoperability** beyond web3

Web3: Potential areas of use for DH

1. **Data Storage** *(Cost & over-centralisation)*
2. **Persistent identification**
3. **Data permanence** *(Maintaining outputs post-grant)*
4. **Interoperability** beyond web3

Data Storage: High costs & over-centralisation

Cloud storage & hosting, industry standard

The Sloane Lab runs on Amazon Web Services

Unmatched benefits: Scalability, flexibility, resources at hand *(Amazon Textract)*

Data Storage: High costs & over-centralisation

But:

(1) Can be prohibitively expensive for small projects

(2) Over 60% of all websites using cloud computing, are served by 3 US-based companies *(Amazon Web Services, Microsoft Azure & Google Cloud)*

In 2021 "*an unexplained configuration error*" (of infrastructure provider Fastly), brought down Gov.uk, the BBC and major commercial websites (Amazon) for 1 hour

Web3 potential for Data Storage challenges

Web3 takes a fundamentally different approach to storage

Web3's storage infrastructure:
Interplanetary File System (IPFS)

Various different decentralised storage providers, but IPFS the undisputed leader

Web3 potential for Data Storage challenges

Versioned, decentralised storage system, which uses content addressing *(Variation of file hash)*

Open-source project, over 4,000 contributors, invented by Julian Bennet of Protocol Labs in 2014

IPFS evolves and connects “proven techniques” from previous peer-to-peer systems, including BitTorrent and Git into a “single cohesive system”

Web3 potential for Data Storage challenges

IPFS is not a blockchain, but a peer-to-peer (P2P) storage system

Blockchain networks serve as incentive layers on top of IPFS, helping to manage storage deals:

- Filecoin *(also run by Protocol Labs)*
- Crust

How to use IPFS?

SLOANE LAB

As in any P2P system, anyone can run a node (an easy-to-use computer app), upload and store files for free *(These will be automatically addressed and federated)*

However, permanence is not guaranteed if our node goes down

On IPFS all files are public

Exploring IPFS for the Sloane Lab

IPFS cannot yet store dynamic websites
(Back-end APIs)

Could be used to store public static datasets

Potential for Knowledge Bases

SLOANE LAB

A fully decentralized triplestore managed via the Ethereum blockchain

Damien GRAUX [a] and Sina MAHMOODI [b]

[a] *Inria, Université Côte d'Azur, CNRS, I3S, France* - damien.graux@inria.fr

[b] *Ethereum Foundation* - sina.mahmoodi@ethereum.org

- **Complete technical "architecture for a linked open data infrastructure**, built on open decentralized technologies"
- Utilises IPFS for data storage and retrieval and also, the Ethereum blockchain network "for **naming**, **versioning** and storing metadata of datasets"
- Supports **SPARQL queries**
- Yet: **Not tested in deployment**; Would require technical expertise also from contributing parties

Data Storage: Cost & Centralisation

Possibilities for decentralised storage have emerged: Primarily for static datasets, but potential also for KBs

Cost:

- 0.002 per GB per Year (over 5x cheaper than AWS)
- Culturally important data: Free

Free Open Data Storage

Incentive layer Filecoin is self-described as a "decentralised storage network designed to store humanity's most important information"

Arweave states it is focused on the "preservation of humanity's most important data"

Common theme emerging across decentralized storage providers

Free Open Data Storage

Filecoin Plus: Sophisticated incentivisation programme offering free storage for projects with valuable Open Access datasets

Currently in use

Filecoin's **Dataset Explorer** hosts over 2,600 TB in Open Access datasets *(Encyclopaedia of DNA Elements, LibriVox Free Audiobook Collection and more)*

Free Open Data Storage: Caveats

Hard to use:

- Platforms meant to make the publication of such data more easily accessible keep changing
- Sloane Lab was offered to do this, through Protocol Labs

Hard to access (Filecoin's Dataset Explorer):

Data Permanence (Post-grant)

What happens once the funding of a fixed-term research project is over?

Web3 provides possibilities for perpetual storage

Incentive layers (blockchains operating on top of IPFS) are able to manage long-term storage deals

Data Permanence (Post-grant)

Discover some real use cases

Uniswap

Check this real-time monitor to see how we help Uniswap decentralized host their Dapp frontend:

81 IPFS Replicas all around the globe

99+ Years guaranteed with **1.5ETH** in payment contract

Guranteed by decentralized incentive protocol on IPFS

Verifiable on IPFS Scan

Risks & Challenges

- Rapidly-evolving, ever-changing landscape
- Permanence not guaranteed *(Even for IPFS)*
- Decentralisation (no oversight): fertile ground for bad actors (9% global phishing threat through IPFS) – But phishing also found on centralised & heavily monitored servers *(Amazon AWS)*
- Broader web3: Poor legal and technical safeguards
- Steep adoption curve

Recommendations

End-user interaction must be web 2.0-like (e.g. fetching a file) *Not creating a blockchain wallet in order to use our application (e.g. to login)*

Monitor developments, there is significant potential *Open data storage, perpetual storage*

To Funders (TaNC): To engage with pioneers from the space (e.g. Protocol Labs) for Knowledge Exchange & to help steer developments *To be beneficial and relevant for Digital Humanities / Research Infrastructures*

SLOANE LAB

Thank you

SloaneLab.org
f.valeonti@ucl.ac.uk

TOWARDS A NATIONAL COLLECTION
UKRI Arts and Humanities Research Council
UCL
The British Museum
NATURAL HISTORY MUSEUM
TECHNISCHE UNIVERSITÄT DARMSTADT