Emerald Journal of Documentation

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

Journal:	*Journal of Documentation*
Manuscript ID	JD-04-2024-0093.R1
Manuscript Type:	Article
Keywords:	Digital Humanities, Blockchain Technology, Decentralised Web (web3), Sloane Lab, Towards a National Collection, Digital Humanities Infrastructures

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, including decentralised storage and persistent identification. It does not provide a holistic overview of all web3 technologies.
Practical implications – The authors identify practical uses of web3 technologies for digital humanities projects, outlining potential applications concerning decentralised storage and persistent identification.
Originality/value – The authors push forward current knowledge and literature on the intersection of web3 and digital humanities, outlining also practical recommendations for scholars, practitioners and funding organisations.

Keywords Digital humanities, blockchain technology, decentralised web (web3), Sloane Lab

1. Introduction

Over the last decade a new form of Internet has been emerging, that of the decentralised web, commonly referred to as *web3*. Leveraging blockchain technology and its ability to enable secure exchanges between strangers without the need for a trusted intermediary, web3 is built upon an infrastructure capable of addressing issues stemming from centralisation. Described as tamper evident and tamper resistant distributed digital ledgers (Yaga et al., 2018), blockchains are increasingly utilised by individuals, organisations and governments seeking to take advantage of their properties of integrity, security, and decentralisation. Indicatively, in 2022 the UK government announced plans to “make Britain a global hub for cryptoasset technology and investment” (HM Treasury, 2022). Although blockchains are predominantly used in finance and supply chain management (Al-Shamsi et al., 2022), decentralised-ledger technologies are adopted across a wide range of sectors including sports, entertainment and culture.

Although technology is a foundational element of digital humanities, web3 remains an understudied field in the domain. Furthermore, in a recent study by Benítez & Romero, the authors identify blockchain as a technology of potentially “great impact” for humanities (Benítez & Romero, 2023, p.103). Aiming to push forward current knowledge, this study examines whether blockchain-based technologies and their capacities for storage, computation and sustainability (i.e. not relying on institutional maintenance and funding), could benefit digital humanities infrastructures, such as digital corpora, archives, platforms and tools. The study identifies four key challenges faced by digital humanities infrastructures, highlighting the limitations of current tools and platforms through the case study of the “Sloane Lab: looking back to build future shared collections”. This 3-year project seeks to reunite the dispersed collections of Sir Hans Sloane and is funded by the Arts and Humanities Research Council’s “Towards a National Collection” programme. The challenges identified are, firstly, the over-centralisation of data storage; secondly, data permanence limitations beyond a project’s funded period; thirdly, limited adoption of persistent identifiers; and, finally, limited data interoperability. The sections that follow are each dedicated to these four challenges in turn, exploring web3 technologies and their potential to help address the aforementioned challenges via decentralised storage, capabilities for self-sustained project funding, decentralised persistent identification and the semantic bridging of data deriving from blockchains with cultural heritage data.

Risks and challenges that may arise from the adoption of blockchain technology in digital humanities are also identified and assessed. Despite the widespread adoption of blockchain applications across industries, decentralised ledgers remain a newly emerging technology plagued by risks, challenges and limitations, which are identified and examined throughout the study. These include application-specific risks (e.g. see section 4.3), as well as web3-wide challenges, such as the lack of legal and technical safeguards, or the high barrier for adoption in terms of technical expertise. Finally, the study also summarises areas of digital humanities for which web3 is deemed unsuitable by design and is not expected to make any substantial contributions in the foreseeable future. This paper aims to inform scholars, practitioners, funding bodies and organisations in the cultural heritage sector and to contribute to an understudied field of research that explores the intersection of web3 and digital humanities. To inform future practice and improve decision making, this paper provides a series of recommendations that are relevant to key groups ranging from individual users and project creators to cultural heritage institutions and funders.

2. Background

2.1. From Digitisation and Aggregation to Digital Humanities Infrastructures

The last two decades have seen a wave of digitisation in the cultural heritage sector. At the national-level, numerous programmes have been implemented, such as the UK’s millennial NOF-Digitise (Nicholson, 2003) and the JISC Content and Digitisation programme (Terras, 2012). At the trans-national level, digitisation across the European Union (EU) has been led by Europeana, a programme funded by the EU that invested 10 million euros annually for a series of years in the digitisation of Europe’s cultural heritage (Europeana, 2014). Following this large-scale digitisation of cultural heritage, domain-specific virtual aggregators were created to facilitate access to dispersed collections within a unified information system. Building a large-scale aggregator itself (europeana.eu), Europeana sought to harvest data whilst leveraging learnings from pre-existing national aggregators on ways to overcome fundamental cross-domain challenges, such as interoperability, fragmentation, copyright and licensing, access and visibility (Purday, 2009). The benefits of aggregators are numerous, as demonstrated by Europeana. Europeana.eu facilitates communication between heritage institutions and audiences, including professional users, whilst helping organisations save significant costs developing similar services and infrastructures themselves (Poort et al., 2013). At the same time, the aggregator provided the general public with unified access to dispersed collections and data.

The emergence of aggregators was followed by the advent of *Linked Data,* a concept that gained in popularity in the heritage sector, aiming to fulfil Tim Berners-Lee's vision for the semantic web (Berners-Lee et al., 2001) as a web of connected data based on common metadata standards (Davis & Heravi, 2021). In 2011, Europeana announced the "*Linked Open Data Pilot*", an "*ongoing effort*" to make the aggregator's metadata "*available as Linked Open Data on the Web*" (Haslhofer & Isaac, 2011) whilst making it possible for data providers (e.g. museums, galleries) to opt for their data to be made available as Linked Data through its platform. In the autumn of 2012 "the full Europeana dataset", then numbering some 36 million objects, was made available as linked open data. In 2017 the OpenAIRE project, which is described as "the point of reference for Open Access in Europe" introduced the OpenAIRE Knowledge Graph, a "rich and up-to-date knowledge graph of research results" (Manghi et al., 2017). Finally, initiatives such as the introduction of Linked Data Fragments by Verborgh et al. (2014) sought to improve scalability of linked data sources.

Although an increasing number of cultural institutions are embracing semantic web technologies, creating projects that leverage linked open data, including the Bibliothèque Nationale de France, the National Library of Spain and the Library of Congress (Candela et al., 2019), the success of the uptake of linked open data in the heritage sector remains unclear, according to interviews conducted by the Sloane Lab (Humbel et al., 2024). Arguably the largest linked open data project concerning cultural heritage yet reaching far beyond it is the free, open data repository of Wikidata (wikidata.org). Unlike its sister project Wikipedia (wikipedia.org), i.e. the volunteer-led encyclopaedia hosted by the non-profit Wikimedia Foundation, Wikidata stores information as linked open data. Heritage institutions and in particular libraries have been using Wikidata as a central platform to link to authority data, as well as to improve local and global metadata quality and processes. Yet, challenges such as loss of control and unclear returns are hindering wider adoption of Wikidata by the heritage sector (Tharani, 2021).

In recent years, focus has shifted towards digital humanities infrastructures. Described as research infrastructures aimed at scholars in the fields of humanities, culture and heritage, digital humanities infrastructures provide technical tools that span from purely digital means (e.g. software platforms and services) to hardware facilities (e.g. high-power computing machinery, digitisation suites, or advanced imaging tools) in order to help them perform their investigation activities (Frosini et al., 2018). Although some relevant projects have been developed outside Europe, such as the LORELEI project for low resource languages funded by the Defence Advanced Research Projects Agency (DARPA) in the US (Strassel and Tracey 2016), the development of digital humanities infrastructures has been funded primarily by the European Commission. Through a plethora of *European Research Infrastructure Consortia* (ERICs), numerous digital humanities infrastructures have been developed that specialise in a range of humanities sub-domains. These include DARIAH for digital humanities, CLARIN for linguistic studies, EHRI for holocaust studies, ARIADNE for archaeology, CENDARI for history and E-RIHS for Heritage Science. The domain-specific approach adopted by digital humanities infrastructures was viewed by some as a barrier to inter-disciplinary research and for that reason projects such as PARTHENOS emerged (Frosini et al., 2018). Described as a "complete technical framework" produced for the "federation of DHIs" (Frosini et al. 2018, p.33), PARTHENOS enables access to resources managed by different digital humanities infrastructures through the use of Virtual

Research Environments; these can be understood as collaborative digital environments that facilitate the integration of resources and tools in order to support research activities (Blanke et al., 2010).

All the same, in the foreseeable future interest in infrastructures is expected to grow. In March of 2023 the European Collaborative Cloud for Cultural Heritage was launched to serve as a "unique infrastructure" aiming to enable "transdisciplinary and large-scale collaboration between cultural heritage professionals" across the European Union (Euro Access, n.d.). In addition, the first strand of work for 2023 for the newly-emerged "common European data space for cultural heritage", which aspires to provide a "framework for supporting and hosting the digital transformation of cultural heritage" (Niccolucci et al., 2022, p.2), is dedicated to "exploring the next generation infrastructure" (Aldana, 2022). In the UK, the "Research Infrastructure for Conservation and Heritage Science" (RICHeS), supported by the Arts and Humanities Research Council, aims to invest more than 59 million pounds to "bring together UK heritage organisations and research organisations to provide them with access to state-of-the-art infrastructure to analyse, digitise and archive national treasures" (National Heritage Science Forum, 2022).

The plethora of digital humanities infrastructures developed to date cater to a wide range of domains. Yet, they require substantial ongoing investment from national and international funding bodies. Beyond the overreliance on centralised funding, such large-scale approaches have also received criticism. Joris Van Zudert argues that the "all-purpose nature" of large digital humanities infrastructures enforces "a generalized strategy aimed at the establishment of standards which is at odds with innovative, explorative research" (Van Zudert 2012, p.184). Claiming that "big, institutionally-based digital infrastructures are a dead end for information technology development and application in the humanities" he dismisses them as "standards-driven", "institutionally-bound", "platforms of exclusiveness" (Van Zudert 2012, p.184). Van Zundert argues instead for a shift towards "an Internet of services where institutions, industry and individuals share computing and storage capacity on an on-demand basis" because that would provide numerous benefits, including cost-efficiency, sustainability, elimination of storage silos and the support of continuous research life cycles (Van Zundert 2012, p.181). A technology that has advanced, optimised and promoted the common sharing of resources concerning computation and data storage over the last decade is that of decentralised ledgers, also referred to as blockchains.

2.2. Emergence of the Decentralised Web

First conceived in 2008 (Nakamoto) by the pseudonymous Satoshi Nakamoto, *blockchain* became widely known as the underlying technical infrastructure of the first cryptocurrency, i.e. *Bitcoin* (Zheng et al. 2018). Enabling transactions between complete strangers without the need for a "trusted third party" (Nakamoto 2008), the blockchain was the technology that managed to achieve consensus and build trust in a peer-to-peer network, eliminating completely the need for any kind of centralised control and oversight. Whereas the Bitcoin blockchain network was solely focused on financial transactions, a few years later Vitalik Buterin explored whether "the blockchain concept can be used for more than just money" (Buterin 2014, p.1). In 2014, Buterin lead the development of Ethereum, the first blockchain network to introduce blockchain-based *smart contracts*. First described as a concept in 1997 by Nick Szabo who also coined the term, smart contracts were first implemented on Ethereum as systems that "automatically move digital assets according to arbitrary pre-specified rules". Ethereum's implementation of smart contracts allowed for all types of computations, opening up possibilities for developing the applications Szabo envisioned in 1997, such as contracts, payments, content rights management and more.

Whereas Bitcoin was the first blockchain network, Ethereum was the first decentralised ledger that allowed for the development of *decentralised applications*, or *DApps*, fostering the advent of the *decentralised web*. Commonly known as *web3*, the decentralised web is often perceived as the next iteration of the web. Underpinned by blockchain technology, web3 is understood as "digital infrastructure whereby protocol-enforced consensus mechanisms facilitate the direct, peer-to-peer, exchange of value between users, removing the need for trusted intermediaries" (Potts & Rennie, 2019, p.93). Web3 has enabled the development of new business models, whilst challenging existing ones. Taking the creative industries as an example, blockchain technology applications claim to be challenging agency-centred business models in favour of artist-centric ones. The affordances of blockchain technology for decentralisation, persistency, auditability and the "automation of payments, licensing, intellectual property management, contracting and governance" (Potts & Rennie, 2019, p.93) have fostered an entire ecosystem of decentralised applications across industries, ranging from supply chain management and logistics, to travel and healthcare.

In the cultural heritage sector, adoption of web3 technologies has been limited, with experimentation so far focusing on Non-Fungible Tokens, or *NFTs*. A new technological medium, described as the digital equivalent of limited editions (Valeonti et al. 2021), and a new art medium in itself (Estorick et al., 2021), NFTs attracted the attention of the cultural sector in March of 2021, when an NFT featuring a digital collage titled "5000 Everydays: The first 5000 Days" by Mike Winklemann was auctioned by Christie's for 69 Million US Dollars. Although the cultural heritage sector tends to be amongst the late adopters of emerging technologies (Valeonti et al. 2021), museums and galleries were amongst the first institutions to experiment with NFTs. This may be attributable to how, early on, NFTs were described as a potential "lifeline" for museums (Ciecko, 2021), emerging right after the pandemic, which had brought unforeseen financial challenges to the sector, forcing institutions into redundancies, closures and financially motivated deaccessioning. Major institutions such as the British Museum, the Uffizi Gallery and the Hermitage Museum explored and experimented with NFTs to raise funds for their collections (Valeonti et al., 2021), significantly promoting the adoption of blockchain technology within the broader cultural sector, albeit not without criticism (Grosvenor, 2022). Beyond culture, as a technological medium that enables digital scarcity, uniqueness and provenance, NFTs are adopted across sectors, ranging from gaming and entertainment (Zhao et al., 2021), to finance and logistics (Al-Shamsi et al., 2022), with the annual trading volume of NFTs in excess of 24 Billion US dollars (Hayward, 2023).

Criticism of decentralised ledger technologies has been mainly focused on their environmental impact; other risks and challenges are covered later in sections 4.3 and 8.3. Environmental criticism stems from the fact the first blockchain ever created, i.e. the Bitcoin blockchain network, is particularly energy intensive. In order to achieve fairness and true randomness amongst peers in a way that cannot be compromised by bad actors, Bitcoin requires excessive computation for transaction validation. As a result, the carbon footprint of the Bitcoin blockchain network is substantial, exceeding that of global gold mining operations, or the footprint of countries as large as Poland and Malaysia (Cambridge Centre for Alternative Finance, 2024). Over the last decade energy-efficient alternatives have emerged and, at present, the vast majority of blockchains beyond Bitcoin are environmentally friendly. These include blockchains stemming from academia, such as Cardano, which was co-developed by the University of Edinburgh's Blockchain Technology Laboratory, and Algorand, which was founded by the MIT Ford Professor of Engineering Silvio Micali. They also include some of the world's largest blockchains by market capitalisation, such as Ethereum, Binance and Solana (CoinGecko, 2024). Ethereum's energy consumption was originally comparable to that of Bitcoin, however, following a major upgrade that took place on the 15th of September of 2022, Ethereum's energy consumption was dropped by 99 per cent according to Alexander Neumueller of the University of Cambridge (Neumueller, 2023). Following the upgrade, which was referred to as "The Merge", Ethereum's energy consumption became comparable to that of standard computing (Ethereum Foundation, 2023).

3. Towards a National Collection, the Sloane Lab and Technical Challenges

Towards a National Collection (TaNC) is a five-year programme funded by the UKRI Arts and Humanities Research Council. With over 18 million pounds of investment, it aims to "*open UK heritage to the world*" and ultimately create "*a unified virtual national collection*" (UK Research and Innovation, 2024). To date, TaNC has funded a series of eight small-scale Foundation Projects to "*lay the foundations for a virtual national collection*" which concluded in 2022 (Towards a National Collection, n.d.), as well as three COVID-19 digital research projects as part of a pandemic-related Urgency Call of TaNC (Towards a National Collection n.d.-b). The majority of the programme's funding, £14.5 million pounds, has been invested in five Discovery Projects for the "*research and development of emerging technologies […] in order to connect the UK's cultural artefacts and historical archives*" (Towards a National Collection, n.d.-c). One of the five Discovery Projects is "*The Sloane Lab: Looking back to build future shared collections*" (Nyhan et al. 2023). It is a collaboration between UCL and TU Darmstadt, the British Museum, the Natural History Museum, proceeding 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. The goal of the Sloane Lab is to reunite the collections of Sir Hans Sloane, which served as one of the founding collections of the British Museum. When the collections of that museum were divided between the present-day British Museum, British Library and London's Natural History Museum, the Sloane collection was dispersed across the three institutions. The Sloane Collection ranges from fossils and vegetable substances to music instruments and printed books. Technical work undertaken by the Sloane Lab has included data aggregation, digitisation, data mobilisation, semantic unification, data modelling, mapping, enrichment and analysis in conjunction with the development of methodological instruments for navigating complex data environments.

Serving as a microcosm of the broader TaNC programme, the Sloane Lab explores, through the virtual reunification of dispersed and disparate collections, solutions for aggregation, interoperability, identification, cross-institution integration and the long-term availability and accessibility of cultural heritage collections. The project employs standard Semantic Web technologies that are proven to facilitate intelligent integration of resources via machine and human-readable representations, promoting interoperability both at a structural and semantics level. Utilising triple-store technology, the Sloane Lab Knowledge Base integrates a wide range of disparate data sources, creating a homogeneous data environment for data interrogation, inference and knowledge discovery. The project integrates both structured and unstructured datasets. The unstructured data (which is subsequently structured and enriched) originates from Sloane's handwritten catalogues containing over 8,300 manuscript pages listing more than 150,000 items, including objects and artifacts, specimens and books. The structured data contain over 20,000 database records of items from the original Sloane catalogues that are now held across the three institutions, the British Library, the British Museum and the National History Museum.

Taking into account the contributions of Web3, which has introduced a range of technologies offering novel infrastructures, processes, and possibilities across key areas of interest for digital humanities, it becomes imperative to assess whether blockchain-based technologies could effectively address some of the challenges encountered by the Sloane Lab and TaNC. We have identified four challenges where existing infrastructure is deficient and web3 technologies hold the potential to provide significant contributions. The areas of digital humanities for which web3 and blockchain technologies are deemed unsuitable and are therefore not expected to have any major impact in the foreseeable future are also examined later in section 7 “Limitations of web3 technologies”.

3.1. Data storage: High costs and over-centralisation

Data produced by projects which are led by a single institution are usually hosted by the organisation’s pre-existing infrastructures, processes and information systems. Yet, digital humanities projects are often collaborative and cross-institutional. That is the case for the Sloane Lab, which is led by a consortium of two universities, UCL and TU Darmstadt, and two museums, London’s Natural History Museum and the British Museum. In these instances data hosting is a challenge, as the responsibility for the long-term maintenance of the project, as well as the value deriving from such an investment, do not rest with a single institution. For the Sloane Lab, all data and web applications are funded, administered and operated by UCL and hosted on the cloud computing provider Amazon Web Services. Utilising cloud computing solutions such as Amazon Web Services and Microsoft Azure is considered the industry standard for web hosting. Yet cloud computing comes with substantial costs for features that may often be unnecessary for research projects. A core advantage of cloud computing, for example, and one of the reasons it is substantially more expensive than traditional web hosting is scalability, i.e. the fact resource allocation can easily be adjusted to address rapid growth to thousands, or millions of concurrent users. Yet, that advantage of cloud computing (the ability to respond and withstand rapid exponential growth), is of little relevance to the majority of research projects which do not tend to have such levels of demographic reach. Beyond cost, centralisation of such services is staggering. Over 60 per cent of all websites using cloud computing are served by only three US-based companies (Amazon Web Services, Microsoft Azure and Google Cloud) with the accumulative market share of this oligopoly increasing year over year (Canalys, 2023). The risks of centralisation are already beginning to show; in June of 2021, a major Internet outage took major publishers and businesses, such as the BBC and Amazon, as well as government websites, including the UK’s gov.uk offline for nearly one hour. The outage was attributed to an “unexplained configuration error” (Sabbagh & Hern, 2021). Other similar events have occurred in recent years, highlighting the lack of resilience that has resulted from over-centralisation, leading some to call for the decentralisation of the web (Maguire, 2021). Examining ways to address the high costs and over-centralisation of web hosting, potential alternatives and solutions emerging from web3 are examined in Section 4.

3.2. Data permanence challenges: Availability of outputs post-grant

The second challenge identified for current digital humanities infrastructures is that of the long-term availability of outputs beyond a project’s funded duration. With the current phase of Towards a National Collection concluding in 2025 and Sloane Lab in its third and final year as a funded research project, the long-term availability of the substantial amount of data and outputs developed arises as a major challenge. The monthly cost of Sloane Lab’s hosting on Amazon Web Services amounts to several hundreds of pounds, making it particularly difficult to maintain the project’s current infrastructure once the funding provisioned to the project for a three-year period has concluded. Section 5 explores ways decentralised storage could support research projects to make their data and outputs available for the long-term, far beyond grant completion. Importantly, leading decentralised storage providers put special emphasis on storing “humanity’s most important information” (Filecoin 2024). That some have introduced and implemented models of sponsored data hosting for such cases suggests that this may be of more than rhetorical and marketing interest to them.

3.3. Limited adoption of persistent identification

A persistent identifier can be described as a long-lasting reference to a digital resource (ORCID, n.d.) such as the Digital Object Identifier (DOI) which is widely used in academia. Persistent identifiers are used broadly to create stable, reliable, trusted links across systems and other institutional boundaries. In the cultural heritage sector, the increasing use of collections as data has led to the exploration of persistent identifiers for heritage entities and data. The TaNC-funded programme “Persistent Identifiers as IRO Infrastructure” explored such use cases, outlining the benefits of persistent identifiers for heritage organisations (Kotarski et al., 2022), with special emphasis on linking dispersed and disparate collections. Stemming from research, persistent identifiers are citation-oriented. However, in the context of cultural heritage, Kotarski et al. argue that linking to entities across the sector is also useful for purposes beyond citation. A report commissioned in 2021 by Jisc (i.e. a major non-profit digital agency supporting tertiary education, research and innovation) suggested that increased adoption of persistent identifiers could save the UK’s research and innovation sector £5.67 million pounds over 5 years, as it would reduce the unnecessary re-keying of data that are readily available through pre-existing persistent identifiers (Brown et al., 2021).

Persistent Identification is a major challenge for data aggregators (Alexiou et al., 2016) and this has also been the case for the Sloane Lab. Adoption of persistent identifiers in cultural heritage remains limited with a recent survey revealing that only 13 per cent of Europeana’s records contain a persistent identifier (Freire 2023). Similarly, in the Sloane Lab, although some of the project’s museum partners have embraced persistent identifiers (e.g. numerous collections of the Natural History Museum are DOI-registered), the vast majority of collections concerning the Sloane Lab have not been registered with a persistent identifier, making cross-institutional data aggregation and handling more challenging. Beyond data aggregators, centralisation has also been identified as an important challenge for persistent identifiers, significantly limiting exposure to research conducted in the global south. Although centralisation of persistent identifiers seemed “unlikely” in the 90s (Gladney, 1998), nowadays, in the context of academic literature, research identifiers are being issued from a very small number of agencies. The effects of centralisation are already beginning to show, with a “large portion of the world’s research” being “invisible to the global north” due to the fact “the rest of the world cannot afford to participate in its indexing, identification, and publishing infrastructure” (ARK Alliance, 2023). A study in 2014 revealed that 84 per cent of peer-reviewed Latin American journals were not indexed by Scopus, or the Web of Science (Babini, 2014). Ways blockchain-based technologies could help advance persistent identification processes and infrastructures are examined in Section 6.

3.4. Interoperability beyond web 2.0

Semantic interoperability is critical for data aggregation platforms and an area of focus for the Sloane Lab. Throughout the project emphasis has been given to semantic interoperability, integration of disparate collections and interchange of heterogeneous cultural heritage information. Yet, despite the widespread use of ontologies across sectors, development and adoption of semantic ontologies for blockchain-derived data is arguably limited. Ontologies are an important layer of the Semantic Web technology stack, as this was envisioned by Berners-Lee et al. (2001). The primary aim of ontologies is to capture domain knowledge with formalised semantics, and, in this way, to enable a shared, unambiguous understanding of domain knowledge. Their main goal is to provide a formal and explicit representation of the knowledge in a specific domain to facilitate knowledge sharing and reuse, reasoning and inferencing, semantic interoperability, integration and standarisation. For cultural heritage specifically, several ontologies have been developed and are widely used across the sector, such as the CIDOC Conceptual Reference Model (CIDOC-CRM), which was also designated as an international standard for the description of cultural heritage resources by the International

Organization for Standardization (ISO 21127:2014). Sloane Lab's data model adopts and extends CIDOC-CRM for facilitating the conceptual modelling and semantic interoperability requirements of the project (Metilli et al., *forthcoming*). The elements of CIDOC-CRM enable the representation of various types of entities, including spatiotemporal entities (e.g. places, events, activities), natural and human-made material things, people and immaterial things (e.g. the contents of a book, or a poem). Yet interoperability of cultural heritage data does not extend beyond web 2.0 with web3-related data remaining semantically separate. That is concerning both data stored in decentralised storage systems and also data stored on blockchains. Considering the long-term interoperability of data produced by the Sloane Lab, Section 7 focuses on ontology development in web3 and possibilities emerging for the interoperability of on-chain (i.e. data deriving from blockchains) and off-chain data (i.e. data residing on traditional infrastructures), such as those produced by Sloane Lab and the four other TaNC Discovery Projects.

4. Exploring Decentralised Storage for the Sloane Lab

In order to explore ways to address the first challenge identified above, i.e. the high-cost and over-centralisation of current web hosting providers, it is necessary to understand the current landscape of decentralised storage. This section examines leading providers, practical options, as well as associated risks and challenges. It concludes with an exploration of possibilities for digital humanities projects and infrastructures, through the case of the Sloane Lab.

4.1. Introduction to the Inter-Planetary File System (IPFS) and the Filecoin Storage Network

Although there are numerous decentralised storage infrastructures available, such as Sia (https://sia.tech), Storj (https://storj.io) and Arweave (https://arweave.org), Filecoin is considered the leading decentralised storage network (Kaur, 2023), storing over 2 Million terabytes of data (Filecoin, 2024b). The storage infrastructure powering the Filecoin Network is the Inter-planetary File System (IPFS), a decentralised storage system that has been widely researched in academic literature (Daniel & Tschorsch, 2022), with some studies examining IPFS itself (Shen et al. 2019, Henningsen et al. 2020), and others proposing solutions that build on top of it, such as "*Healthchain: A blockchain-based privacy preserving scheme for large-scale health data*" by Xu et al. (2019) and "*IoT data privacy via blockchains and IPFS*" by Ali et al. (2017).

A community-driven, open-source project counting more than 4,000 contributors (IPFS, 2024), IPFS was originally conceived in 2014 by Juan Benet. In the whitepaper titled "*IPFS – Content Addressed, Versioned, P2P File System*", Benet defines IPFS as a "peer-to-peer distributed file system that seeks to connect all computing devices with the same system of files" (Benet, 2014). Benet explains that IPFS evolves and connects "proven techniques" from previous peer-to-peer systems, including BitTorrent, i.e. the popular torrent client facilitating peer-to-peer downloads, and also Git, i.e. the widely used version control system into "a single cohesive system". According to Benet, the ultimate goal of IPFS is to develop "new decentralised Internet infrastructure". Describing his motivation, Benet stressed that, whereas the HyperText Transfer Protocol (HTTP) is considered "the most successful distributed system of files ever deployed", it is incapable to meet the challenges of the "new era of data distribution" that we are entering, such as "hosting and distributing petabyte datasets", "computing on large data across organisations", versioning and "preventing accidental disappearance of important files" (Benet, 2014).

IPFS is a peer-to-peer file system, in which any Internet-connected device can participate as a node, without any explicit economic incentives (Henningsen et al., 2020). Similar to proven peer-to-peer filesharing systems (e.g. BitTorrent) each data item is stored by the item's "providers", who are a number of nodes making it available to peers (Henningsen et al., 2020). All data items in IPFS are addressed through immutable, cryptographically generated labels, named “Content Identifiers” (CIDs), which are based on the content’s cryptographic hash signature. It must be noted that CIDs are not the actual file hash, because all contents (e.g. files) on IPFS are divided into blocks, which are organised and verified in a Merkle Directed Acyclic Graph data structure (IPFS, 2024b). CIDs are short in length, typically no longer than 64 characters.

Developed as an “incentive layer” that works on top of IPFS (Protocol Labs, 2017), Filecoin is “a peer-to-peer network that allows anyone to store and retrieve data on the internet” (Filecoin Docs, 2024). Filecoin’s “built-in economic incentives ensure that files are stored and retrieved reliably and continuously” for as long as needed by users. In 2017 Protocol Labs (i.e. a venture founded by IPFS creator Juan Benet) published a whitepaper titled “*Filecoin: A Decentralized Storage Network*” describing Filecoin as a “decentralized storage network that turns cloud storage into an algorithmic market” (Protocol Labs, 2017). In other words, Filecoin provides a storage market, which runs on the homonym blockchain network. Transactions on the Filecoin blockchain network utilise the blockchain’s native cryptocurrency, which is also named Filecoin (FIL), i.e. similar to how Bitcoin (BTC) is the native cryptocurrency of the homonym blockchain. The node operators sustaining the Filecoin blockchain network, also referred to as “miners”, earn Filecoin tokens by providing storage to clients, who, conversely, hire miners to store and distribute their data. IPFS was launched in 2015 as a peer-to-peer storage infrastructure, gaining popularity within the ecosystem of decentralised ledgers (IPFS 2024c).Content delivery and cybersecurity provider, Cloudflare.com, integrated with IPFS in 2018 (Parker), and in 2020 the Filecoin blockchain network was launched.

In March of 2023, the Filecoin Virtual Machine was launched as part of a major upgrade of its blockchain network. Similar to the Ethereum Virtual Machine (explained later in section 7.1), the Filecoin Virtual Machine is a runtime environment for smart contracts and user programmability which runs on the Filecoin network instead (Filecoin Docs 2024b) with the goal to “unlock” the potential of an “open data economy” (Protocol Labs 2024). The Filecoin Virtual Machine was created to enable new use cases to be developed on top of the Filecoin blockchain network, such as perpetual storage and data access control (Filecoin Docs 2024b). The latter which was an exclusive feature of centralised storage solutions prior to the launch of Filecoin Virtual Machine, given that blockchain-based data and data stored on IPFS are always public. Use cases regarding to the former, i.e. perpetual data storage, are described in Section 5.

4.2. Decentralised Storage Options with IPFS

There is a number of ways individuals, organisations and research projects can store their data in a decentralised manner using IPFS. Similar to using peer-to-peer file systems BitTorrent, anyone can download the IFPS software client for Mac, Windows and Linux, which features a self-explanatory user interface, through which users can upload files, including static websites. For all files a Content Identifier is provided as explained above (Section 4.1), that can be shared and used by anyone with an internet browser to access the respective file. By default, IPFS nodes “treat the data they store like a cache” (IPFS, 2024d) unless “pinning” is enabled for files, which can be done through the user interface. Pinning also enables mutability. Content Identifiers are immutable as they derive from the file’s hash. As such, on every file update, the Content Identifier also changes. Pinning provides “mutable pointers” that can be shared and then, in the instance of a file change, these are updated to point to the new Content Identifier.

The IPFS Desktop Client allows anyone to store and share an infinite amount of files free of charge in a transparent and decentralised manner, without any specialist knowledge. Similar to peer-to-peer file sharing, storing and distributing files through the IPFS Desktop Client is free irrespectively of the number of people, who will be requesting and accessing the file. However, if the IPFS node goes offline (i.e. the computer running the IPFS Desktop Client shuts down, or the application is terminated for another reason), there is no guarantee the content stored will be maintained by the network. As such, it is advisable to have the application installed on a number of different computers so that at least one is always online to maintain the files on the network. In order to maintain files on IPFS without using the IPFS Desktop Client, pinning service providers can be utilised. However, these cloud-based services come at a cost that is comparable to conventional, centralised storage providers. Convenient as they are by not requiring any installation and guaranteeing uptime, pinning service providers can be considered as one of the easiest ways to use decentralised storage, yet their subscription model is arguably problematic in the context of research and in particular fixed-term research projects.

For Digital Humanities projects seeking to make the most of IPFS, the use of blockchains that operate on top of its infrastructure would enable more possibilities. Filecoin, mentioned earlier, helps users (i.e. storage clients) and storage providers negotiate and manage storage deals. Importantly, through Filecoin Plus Filecoin opens up possibilities for the storage of open access research data, as examined in “Infrastructure for storing humanity’s most important information” (Section 5.2). Another blockchain network helping users manage storage deals on IPFS is Crust Network. The storage solutions provided by Filecoin and Crust Network come at a lower cost in comparison to centralised providers. Crust Network charges 0.004 US dollars per Gigabyte per year (Crust Network, 2024), Filecoin 0.002 US dollars, whilst Amazon’s centralised AWS S3 servers charges over five times that, beginning at 0.011 US dollars per GB per year (Storage Market, 2024). Arguably, the greatest downside of both Filecoin and Crust Network is that, in order to utilise their solutions, domain-specific, technical expertise is required. Critically, no graphical user interface is provided for managing storage contracts, with the vast majority of operations done through a command-line interface, or other programming interfaces (such through an Application Protocol Interface).

For Digital Humanities projects, the main benefits for using blockchains that manage storage deals on IPFS would be decentralisation, cost-efficiency and the ability to negotiate long-term storage deals, paying upfront for decades-long contracts. Such possibilities are covered in “Perpetual Storage” (Section 5.1). Beyond IPFS, a storage provider with a different approach is Arweave; a blockchain network that stores data on-chain, i.e. on the actual blockchain ledger rather than on a separate peer-to-peer network of nodes as in IPFS, promising “permanent data storage” (Arweave, 2024). However, despite major partnerships, such as with Facebook’s parent company Meta Inc., Arweave’s growth has remained modest (i.e. the total size of data stored at the time of writing is less than 200 Terabytes) (ViewBlock, 2024).

4.3. Risks & Challenges of IPFS

For research projects interested in exploring IPFS, there are several risks and challenges that must be acknowledged and addressed. Firstly, as mentioned above, files stored on IPFS are treated as cache by default. Therefore, it is necessary to ensure that pinning is enabled for all files to guarantee data permanence on the peer-to-peer storage network. Also, if the IPFS Desktop Client is utilised, the data is only pinned on this single node and the application is terminated, the data will eventually be deleted from IPFS. As such, for any project taking advantage of the IPFS Desktop Client it is advised to use it on several computers with the largest uptime possible to guarantee there will always be at least one node with all files pinned online. Secondly, network latency must be taken into consideration. The time required for file retrieval varies significantly, depending on the network’s state and on the popularity of the data (Kaleido, 2023), which is similar to other peer-to-peer file sharing systems.

Arguably the greatest challenge concerning the adoption of decentralised storage solutions is the technical barrier affecting both project creators, as well as end-users. Project developers must possess specialist knowledge and they must also invest a substantial amount of time in order to use decentralised storage effectively and securely. Content stored on IPFS is generally accessible through Internet browsers (using IPFS gateways such as https://ipfs.io). Yet, in some instances, the technical barrier also affects end-users. The Filecoin Dataset Explorer is a large repository of Open Access datasets (Section 5.2). However, the only way to download a dataset is through a command line interface, which is alienating for most users. More broadly, IPFS has also received criticism for its poor content moderation. Due to its decentralised nature, IPFS is fertile ground for bad actors seeking to host fraudulent content. As of October 2022, IPFS links "represented 9% of the global phishing threat" (Pernet et al. 2023), enticing unsuspected users to disclose their credentials to criminals. Although that is indeed significant, it must be noted that phishing attacks have also been documented to be "hiding" behind centralised storage solutions, where monitoring is possible and control is much greater (Burt, 2022).

4.4. Exploring the Potential of Decentralised Storage for the Sloane Lab

The Sloane Lab Knowledge Base uses the GraphDB triple store hosted on Amazon AWS. Users can access and query the contents of the Knowledge Base through the Metaphactory data framework and visual interface, which are also hosted on the same cloud service provider. As mentioned previously, hosting dynamic websites on decentralised storage infrastructures is not yet possible, and this also applies to platforms such as Metaphactory. However, studies have demonstrated that there is potential for using decentralised infrastructures to host decentralised knowledge bases. Arguing that "decentralized file systems could be used to support better availability and performance" of Linked Open Data (LOD), whilst "preserving [their] principles", in 2016, Sicilia et al. presented a "first prototype design of LOD over the Interplanetary File System" (Sicilia et al. 2016). A year later, Cochez et al. (2017) proposed a "deployment of Linked Data" and prototype-based Knowledge Bases on top of IPFS, which, according to the authors, "has several useful features matching the needs for knowledge representation based on prototypes" (Cochez et al. 2017, p.1997). The authors argue that the deployment of linked data in a distributed environment has the potential "to deliver the original idea of the Semantic Web as a web of data" (Cochez et al. 2017, p.2005).

Whereas the approaches of Sicilia et al. and Cochez et al. were largely theoretical, in 2021 Damien Graux of Universite de Cote d'Azur and Sina Mahmoodi of the Ethereum Foundationpresented a complete technical "architecture for a linked open data infrastructure, built on open decentralized technologies" (Graux & Mahmoodi 2021, p.183). The implementation of Graux and Mahmoodi is published on GitHub and utilises IPFS for data storage and retrieval, as well as the Ethereum blockchain network "for naming, versioning and storing metadata of datasets". Critically, the solution proposed by the authors includes "an indexing scheme suitable for linked data and a mechanism for retrieval of data by performing triple pattern or SPARQL queries" (Graux & Mahmoodi 2021, p.183-184). It also outlines "how smart contracts can be employed to provide a persistent identifier for data objects stored on IPFS, to describe and version datasets, to control write access and to ensure source of provenance".

Although Graux & Mahmoodi provide the source code for developing a knowledge base with version control, custom permissions and the ability to retrieve data using SPARQL queries, implementing and operating their solution would require expert knowledge of engineering on decentralised ledgers and familiarity with IPFS. Beyond development, it would also require the parties making changes to the data (e.g. in the case of the Sloane Lab that would be the data providers) to be familiar with decentralised environments and utilising the Ethereum blockchain. Finally, every interaction with Ethereum incurs transaction fees, the so-called "gas fees", meaning that potential transaction fees should also be taken into consideration. Taking all of this into consideration, decentralising the Sloane Lab Knowledge Base would be a separate project in itself, which would involve not only its technical implementation but also the training of the parties involved to utilise and update the knowledge base. Considering that a decentralised triple store is not currently an option for the Sloane Lab, decentralised storage could instead be utilised to host a static export of the data in formats that would be useful for future scholars and practitioners. The challenges to be addressed in that case would be to guarantee seamless retrievability through a user-friendly interface and to consider any related costs; these are examined in the next section.

5. Possibilities for the Long-term Availability of Project Outputs

5.1. Perpetual storage

As mentioned earlier (Section 4.2), there are numerous ways to leverage decentralised storage infrastructure. Cost and complexity vary between solutions and tend to be inversely related; the lower the complexity the higher the cost. As it is often the case for research grants, Sloane Lab is funded for a limited period of time and because of that solutions that require ongoing charges are excluded, such as using IPFS pinning providers (Section 4.2.2). However, beyond subscription-based solutions, decentralised storage infrastructures offer possibilities for securing long-term data storage produced by limited-duration research grants. Although there is limited information available on long-term storage deals and their technicalities, there are two cases that demonstrate the potential of such solutions, which, importantly, concern live deployments.

One of the leading decentralised cryptocurrency exchanges Uniswap (https://uniswap.com) has been using Crust Network to help host the front-end of their DApp on IPFS with “99+ years guaranteed with 1.5 ETH in payment contract” according to the storage provider (Crust Cloud, 2024). Similarly, open-source liquidity protocol AAVE (https://aave.com) host their DApp front-end on IPFS with similar terms, i.e. “99+ Years guaranteed with 1.5 ETH in payment contract”, according to Crust Network. Even if we were to take the highest price of ETH in history, 4,891.70 US Dollars (Coinbase, 2024), the payment contract for either of those projects would not exceed 7,337.55 US Dollars or 5,780.33 British Pounds for nearly a century. However, the process of hosting Uniswap and AAVE on IPFS is elaborate (Crust Network, 2021; Crust Network, 2022). Critically, both AAVE and IPFS use Crust Network alongside other pinning providers, to maximise replications across the IPFS network to reduce latency, yet the costs mentioned above offer 75 replications (i.e. different storage providers offering the same content) for Uniswap and 85 for AAVE as of the time of writing (Crust Cloud, 2024).

Beyond paid options, the infrastructure of Filecoin Plusallows for the free storage of projects. Although technical expertise is required in order to take advantage of these options, the infrastructure has been developed to the extent thatit is already being utilised and therefore it can be leveraged by any project seeking to invest the time to explore it.

5.2. Infrastructure for Storing

For Digital Humanities projects and research projects with Open Access datasets it is of interest that decentralised storage providers have been developing infrastructure for the storage of such datasets free of charge. Arweave was first to announce initiatives such as the decentralised public library, which aspired to defeat “data death”, protecting data against a number of risks, including natural phenomena and censorship (Arweave, 2018). Although Arweave’s decentralised public library was never materialised, Filecoin launched Filecoin Plus in 2019, a system whose goal was to increase its ability to become the leading decentralised storage network (Filecoin Docs, 2024c). Filecoin Plus achieves that through a sophisticated incentivisation model, which orchestrates the interaction between “notaries, [storage] clients, and storage providers”. Additionally, Filecoin’s Dataset Explorer allows Internet users to “discover and explore Open Access datasets stored around the world on the

Filecoin network” (Filecoin, 2024c). However, there is major drawback with regards to the Dataset Explorer; in order to download any dataset, one must install a retrieval client and use a command line interface. Given the rapid pace decentralised ledger technologies are developing, thiscould be considered a transient challenge of a nascent technology. From the perspective of Digital Humanities, it is encouraging to witness leading decentralised storage providers investing in the development of sophisticated incentivisation models to provide free storage to publishers of Open Access datasets.

6. Persistent Identification possibilities in web3

The centralisation of persistent identifiers and the sometimes prohibitive costs that are often associated with them for the Global South which in turn limits research exposure to the Global North, deem it necessary to explore alternatives. One of the most widely used persistent identifiers in the cultural heritage sector is the Archival Research Key or ARK (Kunze, 2003; Kelly et al., 2021; Laddusaw, Sare, & Buckner, 2017). The ARK identifier was introduced in 2001 and it was amongst the first to feature free, decentralised registration and management of persistent identifiers. Yet, at present “known ARK implementations are mostly centralised” (Segundo et al., 2023). In an effort to make ARK truly decentralised Segundo et al. developed a decentralised implementation of the identifier named “dARK”. In the whitepaper titled “dARK: A decentralized blockchain implementation of ARK Persistent Identifiers” they present their implementation, which provides a “decentralized procedure for assigning ARK persistent identifiers” whilst providing an open, “non-centralized permissioned public identifier factory and resolver service” (Segundo et al. 2023, p.1). Segundo et al.’s were aiming to pilot dARK in Brazil because ultimately their goal is to “contribute to the advance of Open Science, mainly in the global south, where just a small portion of scholarly communication has access to PIDs” (Segundo et al. 2023, p.20).

Instead of adapting a pre-existing identifier to leverage blockchain technology, Sicilia et al. sought to leverage the inherent ability of decentralised storage infrastructures for content addressing and more specifically for the “decentralised maintenance of digital objects retrievable by content hashing” (Sicilia et al. 2019, p.126). Arguing that existing systems for persistent identification “still depend on centralized services which may either become unavailable or cease to behave or be managed as expected”, Sicilia et al. showed that decentralised technologies could serve as “infrastructure for the design of a trustless PID system” (Sicilia et al. 2019, p. 123), i.e. one that does not rely on a trusted centralised, intermediary (e.g. for resolution). The authors introduce a novel persistent identification system that uses IPFS “as supporting infrastructure” along with “a minimal set of conventions” that could serve as “an alternate mechanism for storage of byte streams in [major] digital repositories as DSpace or CKAN”. The ability to decentralise the resolution of persistent identifiers enables the “decoupling [of] PID systems from trusted parties” making them more robust and tamper-proof, according to the authors (Sicilia et al. 2019, p. 129). Although both solutions are not live deployments from which the academic community and the cultural heritage sector could benefit, they demonstrate possibilities that are fundamentally different from current approaches, which have the potential to enable truly decentralised issuance and resolution of persistent identifiers.

7. Bridging blockchain-related data with CIDOC-CRM

Extending digital humanities to the realm of web3 semantically and in an interoperable manner would require an ontology-based data model that integrates concepts from both domains. However, despite the wealth of research on ontologies and their widespread use for data modelling in Cultural Heritage and Digital Humanities projects, when it comes to web3 (data stored on blockchain ledgers and decentralised storage infrastructures) the use of ontologies remains limited with only few ontologies specifically designed for such data. BLONDiE is an ontology developed in 2020 by researchers at the

Universidad Nacional de San Antonio Abad del Cusco in Peru. The ontology's aim is to semantically describe the native structures and related information of the Bitcoin blockchain and the Ethereum blockchain (Ugarte-Rojas and Chullo-Llave, 2020). The focus of the ontology is highly technical, concerning the core functionality of decentralised ledgers. Similarly, the Ethereum Ontology (EthOn) covers technical concepts relating to blockchain, such as blocks, accounts and transactions (ConsenSys, n.d.). In 2022, Besançon et al. extended the EthOn ontology to capture "key concepts related to DApps development" and their interrelations (Besançon et al., 2022, p.49905), developing the "DApps ontology" which they validated within the industrial video game Light Trail Rush and its NFT management DApp (Besançon et al., 2022). All these efforts are significant towards the general goal of making blockchain applications interoperable and their data more findable, accessible and reusable. However, due to their focus on technical aspects and the lack of links to standard ontologies from other domains, they cannot support information discovery and exchange across different domains, which is essential for Digital Humanities.

A recent attempt towards semantically linking blockchain and Digital Humanities concepts is the Blockchain Technology for Cultural Heritage (BlockTeCH) ontology. Developed by Thomas Crane (2022) as part of his MSc dissertation at University College London, BlockTeCH integrates elements of CIDOC-CRM, BLONDiE and EthOn, aiming to capture technical aspects and, more importantly, "*the history and cultural heritage of blockchain*" as Crane explained. BlockTeCH was implemented as a specialisation of CIDOC-CRM. It uses classes and properties of CIDOC-CRM to model amongst other concepts, transactions (as instances of *E8 Acquisition*), bitcoin mining (using *E65 Creation*) and legal rights (using *E30 Right*). BlockTeCH also introduces new classes and properties in order to model concepts that cannot be directly captured by CIDOC-CRM, such as *E32a Blockchain Protocol* (as a subclass of *E31 Document*), *E29a Machine Language Instructions* (as a subclass of *E73 Information Object*), *E7a Code Execution (*as a subclass of *E7 Activity*) and *E73a Digital Asset* (as a subclass of *E90 Symbolic Object*). The classes and properties of BLONDiE and EthOn are integrated into BlockTeCH to enable a more detailed description of the technical aspects of blockchains. All classes of BLONDiE and most classes of EthOn are defined as subclasses of *E29a Machine Language Instructions*, while most properties are defined as subproperties of *P165a encodes*. Having been implemented as a specialisation of CIDOC-CRM, BlockTeCH is compatible with any other extension of CIDOC-CRM, as well as with any knowledge base constructed using CIDOC-CRM such as the Sloane Lab Knowledge Base. Ontologies such as BlockTeCH could serve as the semantic middleware for integrating data available in knowledge bases like Sloane Lab's with the vast amounts of data available in decentralised infrastructures, contributing towards the decentralisation of digital humanities.

8. Limitations of web3 technologies in the context of Digital Humanities

Whereas blockchain-based technologies open new possibilities for decentralised data storage and persistent identification using content identifiers and offers novel capabilities concerning the long-term availability of project outputs with perpetual storage, there are core aspects of the Sloane Lab and digital humanities infrastructures more broadly where web3 technologies are unsuitable for making any important contribution in the foreseeable future. Below we examine two such functions, i.e. data analysis and digitisation, and explain for each one the limitations of web3.

8.1. Data Analysis

With the introduction of smart contracts Ethereum also introduced the Ethereum Virtual Machine (EVM), which is a runtime environment on which smart contracts are deployed (Khoury et al 2018). Although implementations differ significantly, most, if not all, blockchains that succeeded Ethereum offer smart contract functionality and therefore their own infrastructure for an EVM-equivalent. In theory, it can be argued that blockchains with smart contract functionality like Ethereum are able to process and analyse data because they offer smart contracts (in essence, software programmes) and a runtime environment. In practical terms, however, that is not possible for two reasons. Firstly, smart contract deployment on EVM occurs transaction fees, which are directly related to computational effort, as well as storage allocation. As such, data analysis in a decentralised environment would be costly. Additionally, smart contracts are executed in a siloed fashion, only having access to on-chain data during runtime. In order to allow smart contracts to access data that is stored outside a given blockchain oracles must be utilised, which are "data feeds that bring data from off the blockchain (off-chain) data sources […] on the blockchain (on-chain) for smart contracts to use" (Ethereum Foundation, 2024). As a result, even simple processes such as obtaining data from public Application Protocol Interfaces (APIs) are complicated in a decentralised environment.

Whilst the requirement of oracles for off-chain data access adds an additional layer of complexity, what makes operations like data analysis prohibitive in decentralised environments is the cost of on-chain computation. In order to counter this challenge, DApps often involve traditional server-side applications, which interoperate with the blockchain. Operating as common web applications otherwise, these applications only interact with the blockchain when necessary, i.e. to perform on-chain transactions, or to extract information from the blockchain. Even for DApps, where most operations, logic and computation take place on decentralised ledgers using smart contracts, it is often the case to partly use traditional servers due to the fact dynamic websites cannot yet be hosted on decentralised storage infrastructures. For that reason, solutions that bridge blockchain applications with traditional infrastructures have been emerging (Avrilionis & Hardjono, 2021). As such, at present for several operations including high computation and the hosting of dynamic websites, web3 cannot compete with traditional infrastructures.

8.2. Digitisation

Digitisation is considered a core function of digital humanities. However, the hardware infrastructure web3, be it blockchains such as Ethereum, or decentralised storage systems such as IPFS, is limited to a network of computers across the globe. As such, decentralised ledger technologies cannot directly support activities that require custom machinery. Yet, it is worth noting that although the role of blockchains could be limited for a digital humanities infrastructure such as E-RIHS, it could instead be transformative for PARTHENOS (both projects mentioned earlier in section 2.1). As a federation that aims to enables "transparent access to resources managed by different [Digital Humanities Infrastructures]", PARTHENOS seeks to create "an homogenous information space where all resources (data, services and tools) of the different DHIs are described" and shared efficiently (Bardi & Frosini, 2017). Federation is described as a solution that helps orchestrate services and resources in multi-domain scenarios. In dynamic environments, federation poses numerous challenges such as availability of resources, security and privacy (Antevski & Bernardos, 2022). In the publication "*Federation in dynamic environments: Can Blockchain be the solution?*" the authors propose and validate a high-level design of a blockchain application for federation, demonstrating that the application of decentralised ledger technologies can "successfully tackle most of the challenges" of federated dynamic environments. Nevertheless, and in comparison to current infrastructure, beyond a more efficient federation, web3 technologies have little to contribute to any aspect of digital humanities involving custom hardware.

9. Observations

Based on the above investigation of web3 technologies through the lens of digital humanities, the following observations can be made, which can be broadly organised into three categories: opportunities, barriers to adoption and risks.

9.1. Opportunities: Decentralisation and Open Access

As demonstrated in previous sections, web3 has introduced an array of novel tools and processes that are of interest for a range of different aspects of digital humanities. These technological solutions are fundamentally different to the Internet as we know it as they are premised on decentralisation, showing significant potential for digital humanities infrastructures. The ability to store vast amounts of data without relying on institutional funding, or without relying on the centralised servers of a commercial firm has the potential to create new norms on data publishing and dissemination. Additionally, the ability to resolve persistent identifiers without the need for a centralised entity, enables the creation and adoption of a truly decentralised persistent identifier.

Beyond the purely technical innovations introduced by web3 other opportunities are presented, especially for digital humanities projects working with Open Access data. Materialising a vision shared across different decentralised storage providers, i.e. to store “humanity’s most important information” (Filecoin 2024), Filecoin Plus was developed asan elaborate system that orchestrates data providers, storage providers and notaries to allocate free data storage to projects that publish open access data. A manifestation of Filecoin Plus is the Dataset Explorer, which at present serves more as an archiving facility rather than actual storage as technical expertise is required to retrieve data from it (Section 4.3). Nevertheless, with web3 technologies evolving at a rapid pace, it could be considered a matter of time until both decentralised data storage and data retrieval become more accessible to data publishers and end-users respectively.

9.2. Barriers to Adoption

The key barrier to adoption by the cultural heritage sector and Digital Humanities more broadly emerging across all use cases examined is the highly specialised technical expertise required in order to make the most of these technologies. The decentralised technologies examined in this study are at varying levels of development and maturity, with some of them being at more theoretical, primitive stages and others, such as decentralised storage, already widely used storing millions of terabytes of data. In order to take advantage of any of these solutions familiarity with using blockchains is required, as well as with decentralised storage systems such as IPFS. Technical expertise in itself is not sufficient, as decentralised ledgers are fundamentally different in comparison to common Internet technologies. Finally, for Digital Humanities in particular, although technical solutions have been developed which can be beneficial for the domain, such as a dARK for persistent identification and BlockTeCH for the semantic bridging of on-chain and off-chain data, these are still in primitive stages of development, hindering possibilities of experimentation with web3.

9.3. Risks & Challenges

Before any solution is adopted beyond research and experimentation risks and challenges must be investigated and addressed. Firstly, there are project-specific risks, such as those stated for IPFS (Section 4.3). Secondly, there are broader challenges when adopting decentralised ledger technologies. Because of the irreversible nature of blockchains, he lack of basic technical safeguards, such as “Undo” and “Reset your password” (Valeonti, 2022), comprise a major disadvantage when operating in a fully decentralised environment. Indicatively, in 2021 the ZKM Centre for Art and Media in Karlsruhe, Germany accidentally lost two Crypto Punk NFTs it had in its possession since 2017. These were valued at 60 ETH (i.e. more than 100,000 US dollars) each due to a “stupid copy-paste error” (Batycka, 2020). The lack of regulation is also another challenge in web3 as is the lack of legal safeguards (Flick, 2022). Additionally, decentralisation entails less oversight, if any. As mentioned in section 4.3, due to its decentralised nature IPFS is fertile ground for hosting fraudulent, hateful and illegal content, more so than traditional servers where take down notice and takedown procedures are swift. It must be noted that even in major centralised providers, such as Amazon’s AWS, fraudulent content has also been found (Section 4.3). Nevertheless, the technical landscape is quickly adjusting. For example, whereas copyright infringement had been a major challenge of NFT platforms, now the UK’s Design and Artists Copyright Society (DACS) states that “most NFT marketplace platforms [operate] a notice and takedown procedure for requesting removal of infringing content from their platforms” (DACS, n.d.). As individual use cases vary greatly, the assessment of any risks should focus on each use case and the technical solutions adopted. Beyond that it should also concern broader risks and challenges of web3.

10. Recommendations

Given the breadth and the novelty of technologies web3 introduces, with potential to unlock new possibilities; it is recommended to digital humanities scholars, practitioners, institutions and funding bodies to closely monitor developments in the field. Although the barriers to adoption remain significant, including the highly specialised, domain-specific technical expertise required and the fact that solutions specifically aimed at digital humanities are at their primitive stages, the pace of development of decentralised ledger technologies is rapid. The risks identified in this study must be acknowledged, investigated and individually addressed before adopting web3 technologies beyond experimentation. These risks concern, firstly, the given use case and the technical solution adopted, and, secondly, the challenges of decentralised ledger technologies more broadly as explained in the previous section. Conducting this study a set of practical recommendations can be made and these are outlined next.

Firstly, as mentioned above, decentralised ledgers are lacking the sophisticated technical safeguards of the Internet as we know it where actions can be undone and lost accounts can be retrieved. In order to overcome that risk, it is advisable for project developers to avoid building DApps that require inexperienced end-users to interact with the blockchain directly (e.g. requiring them to create a wallet in order to interact with the application). DApps aimed at experienced blockchain users should not be of concern. For Digital Humanities and also for research projects, solutions that could be pursued are those that limit web3 use to the application’s internal operations, which do not involve the end user. Such an example could be a Digital Humanities project that only uses web3 technologies for archiving or storage, e.g. storing some of its data in a decentralised infrastructure, such as IPFS, which is otherwise accessible through a web browser. Such a solution would leverage decentralisation and decentralised storage in particular, without alienating end users or exposing them to any risk.

For researchers, the dearth of knowledge on the subject provides fertile ground for making an impact in a rapidly expanding field of research. For practitioners, experimentation to gain familiarity and a deeper understanding of decentralised ledger technologies could be beneficial in order to be better equipped to follow developments and identify potential opportunities in the future. For funders, investment on the intersection of web3 and Digital Humanities would foster innovation and help develop truly decentralised solutions and standards. For TaNC, although any deployment-level utilisation of web3 technologies stretches beyond the scope of the programme, in a future programme it could be worth exploring the capabilities of decentralised storage and whether that could be leveraged for the benefit of the national collection. That could be achieved by establishing partnerships and collaborations with stakeholders from within the industry, and in particular with the organisations that lead the efforts for free Open Access data storage (Section 5.2).

10. Conclusion

Aiming to fill the dearth of knowledge on the intersection of web3 and digital humanities, this study provided an overview of ways decentralised technologies could be of benefit to digital humanities infrastructures whilst highlighting associated risks. Driven by the needs of the Sloane Lab, a TaNC-funded project bringing together the collections of Sir Hans Sloane into one infrastructure, this study explored ways web3 could be of benefit for digital humanities infrastructures, examining decentralised storage, persistent identification, semantic bridging of blockchain data with cultural heritage data, as well as options for the long-term storage of outputs produced in limited duration research projects. The study presented three observations, which identify, firstly, the opportunities for decentralisation and open access data storage, secondly, the barriers to adoption and, thirdly, associated risks and challenges. The study concludes with recommendations for digital humanities scholars, practitioners and funders. With interest in web3 growing and scholars and practitioners increasingly seeking to understand the opportunities and risks of blockchain-related technologies, this paper seeks to help scope out future research and practice on the cross-section of decentralised ledger technologies and digital humanities more effectively.

Bibliography

Aldana, L. (2022, November 30). Europeana Initiative begins building the common European data space for cultural heritage. *Europeana PRO*. https://pro.europeana.eu/post/europeana-begins-building-the-common-european-data-space-for-cultural-heritage

Alexiou, G., Meimaris, M., & Papastefanatos, G. (2016). Enabling persistent identification of groups of duplicates in data aggregators. In *2016 IEEE 32nd International Conference on Data Engineering Workshops (ICDEW)* (pp. 124-126). Helsinki, Finland. https://doi.org/10.1109/ICDEW.2016.7495630

AlShamsi, M., Al-Emran, M., & Shaalan, K. (2022). A Systematic Review on Blockchain Adoption. *Applied Sciences*, *12*(9). https://doi.org/10.3390/app12094245

Antevski, K., & Bernardos, C. J. (2022). Federation in Dynamic Environments: Can Blockchain Be the Solution? *IEEE Communications Magazine, 60*(2), 32–38. https://doi.org/10.1109/MCOM.001.2100585

ARK Alliance. (2023). The ARK origin story. Retrieved April 11, 2024, from https://arks.org/about/the-ark-origin-story/

Arweave. (2018, September 28). Announcing the Decentralised Public Library. Medium. https://arweave.medium.com/announcing-the-decentralised-public-library-cc3dd0235d3

Arweave. (2024). Use Arweave. Retrieved April 11, 2024, from https://www.arweave.org/use

Arweave. (2024b). Use Arweave. Retrieved April 11, 2024, from https://www.arweave.org/

Avrilionis, D., & Hardjono, T. (2021). Towards Blockchain-enabled Open Architectures for Scalable Digital Asset Platforms. *arXiv preprint arXiv:2110.12553*.

Babini, D. (2014). APCs: The new enclosure to knowledge [Slideshare slides]. SlideShare. Retrieved from https://www.slideshare.net/Babini/coasp2014-dominique-babini-clacso

Bardi, A., & Frosini, L. (2017). building a federation of Digital Humanities Infrastructures. *Digital Humanities*, *28*.

Batycka, D. (2022, January 21). A German Museum Has Accidentally Lost Access to Two Highly Valuable NFTs. *The Art Newspaper*. Retrieved from https://www.theartnewspaper.com/2022/01/21/a-german-museum-has-accidentally-lost-access-to-two-highly-valuable-nfts

Benet, J. (2014). Ipfs-content addressed, versioned, p2p file system. *arXiv preprint arXiv:1407.3561*.

Benítez, F. & Romero, E. (2024). What is Blockchain and How can it Help the Humanities?. In A. Gallego Cuiñas & D. Torres-Salinas (Ed.), *Humanities and Big Data in Ibero-America: Theory, methodology and practical applications* (pp. 93-108). Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110753523-007

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). Web Semantic. *Scientific American*, *284*(5), 34-43.

Besancon, L., Da Silva, C. F., Ghodous, P., & Gelas, J. P. (2022). A blockchain ontology for DApps development. *IEEE Access*, *10*, 49905-49933.

Blanke, T., Candela, L., Hedges, M., Priddy, M., & Simeoni, F. (2010). Deploying general-purpose virtual research environments for humanities research. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *368*(1925), 3813–3828. https://doi.org/10.1098/rsta.2010.0167

Brown, J., Jones, P., Meadows, A., Murphy, F., & Clayton, P. (2021). *UK PID Consortium: Cost-Benefit Analysis*. Zenodo. https://doi.org/10.5281/zenodo.4772627

Burt, J. (2022, August 22). Hiding a phishing attack behind the AWS cloud. The Register. https://www.theregister.com/2022/08/22/aws_cloud_phishing/

Buterin, V. (2014). *Ethereum whitepaper*. Ethereum. Retrieved from https://ethereum.org/content/whitepaper/whitepaper-pdf/Ethereum_Whitepaper_-_Buterin_2014.pdf

Cambridge Centre for Alternative Finance. (n.d.). Cambridge Bitcoin Electricity Consumption Index. *University of Cambridge Judge Business School*. Retrieved April 10, 2024, from https://ccaf.io/cbnsi/cbeci

Canalys. (2023, February 8). Worldwide cloud service spend to grow by 23% in 2023. *Canalys Newsroom*. https://www.canalys.com/newsroom/global-cloud-services-Q4-2022

Candela, G., Escobar, P., Carrasco, R. C., & Marco-Such, M. (2019). A linked open data framework to enhance the discoverability and impact of cultural heritage. *Journal of Information Science, 45*(6), 756-766. https://doi.org/10.1177/0165551518812658

Ciecko, B. (2021, March 10). The surging demand for digital collectibles could offer a lifeline for cash-strapped museums—Here's how. *Artnet News*. https://news.artnet.com/art-world-archives/op-ed-digital-collectables-museums-1950808

Cochez, M., Hüser, D., & Decker, S. (2017, June). The future of the semantic web: Prototypes on a global distributed filesystem. In *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)* (pp. 1997-2006). IEEE.

Coinbase. (2024). Ethereum price. Retrieved April 11, 2024, from https://www.coinbase.com/en-gb/price/ethereum

CoinGecko. (2004). Layer 1. Retrieved April 11, 2024, from https://www.coingecko.com/en/categories/layer-1

ConsenSys. (n.d.). EthOn. GitHub. Retrieved April 11, 2024, from https://github.com/Consensys/EthOn

Crane, T. (2022). *BlockTeCH: An Ontology for the Cultural Heritage of Blockchain Technologies* [Master's thesis, University College London]. University College London.

Crust Cloud. (2024). Home. Retrieved April 11, 2024, from https://crustcloud.io

Crust Network. (2021, March 31). Decentralized Uniswap Interface Hosting on IPFS. Medium. https://medium.com/crustnetwork/decentralized-uniswap-interface-hosting-on-ipfs-18a78d1209ac

Crust Network. (2022, August 17). Decentralized Aave Interface Hosting on IPFS. Medium. https://medium.com/crustnetwork/decentralized-aave-interface-hosting-on-ipfs-ed81cff1d3e2

Crust Network. (2024). Home. Retrieved April 11, 2024, from https://crust.network

Davis, E., & Heravi, B. (2021). Linked Data and Cultural Heritage: A Systematic Review of Participation, Collaboration, and Motivation. *Journal on Computing and Cultural Heritage*, *14*(2). https://doi.org/10.1145/3429458

Design and Artists Copyright Society (DACS). (n.d.). Copyright uncovered: A beginner's guide to NFTs. Retrieved from https://www.dacs.org.uk/news-events/copyright-uncovered-a-beginners-guide-to-nfts

Estorick, A., Waters, K., & Diamond, C. (2021, April 10). In search of an aesthetics of crypto art. *Artnome*. https://www.artnome.com/news/2021/4/10/in-search-of-an-aesthetics-of-crypto-art

Ethereum Foundation. (2023, October 24). Energy consumption. Retrieved April 11, 2024, from https://ethereum.org/en/energy-consumption/

Ethereum Foundation. (2024). Oracles. Ethereum Developers Documentation. Retrieved from https://ethereum.org/en/developers/docs/oracles/

Europeana. (2014). *Europeana Strategy 2015-2020*. https://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20Strategy%202020.pdf

Euro Access. (n.d.). Launch of the European Collaborative Cloud for Cultural Heritage. Euro Access. Retrieved July 25, 2024, from https://www.euro-access.eu/en/news/34/Launch-of-the-European-Collaborative-Cloud-for-Cultural-Heritage

Filecoin. (2024). Home. Retrieved April 1, 2024, from https://filecoin.io

Filecoin. (2024b, April 11). Filecoin Storage Stats. Retrieved April 11, 2024, from https://storage.filecoin.io

Filecoin. (2024c). Dataset Explorer. Retrieved August 18, 2024, from https://datasets.filecoin.io

Filecoin Docs. (2024). Welcome to Filecoin Docs. Retrieved April 11, 2024, from https://docs.filecoin.io

Filecoin Docs. (2024b). The Filecoin Virtual Machine. Retrieved April 11, 2024, from https://docs.filecoin.io/smart-contracts/fundamentals/the-fvm

Filecoin Docs. (2024c). Filecoin plus. Retrieved April 11, 2024, from https://docs.filecoin.io/basics/how-storage-works/filecoin-plus

Flick, C. (2022). A critical professional ethical analysis of Non-Fungible Tokens (NFTs). *Journal of Responsible Technology, 12*, 100054. https://doi.org/10.1016/j.jrt.2022.100054

Freire, N., Manguinhas, H., Isaac, A., & Charles, V. (2023). Persistent Identifier Usage by Cultural Heritage Institutions: A Study on the Europeana.eu Dataset. In O. Alonso, H. Cousijn, G. Silvello, M. Marrero, L. Teixeira Lopes, & S. Marchesin (Eds.), *Linking Theory and Practice of Digital Libraries. TPDL 2023* (pp. 124-126). Lecture Notes in Computer Science, Vol. 14241. Springer, Cham. https://doi.org/10.1007/978-3-031-43849-3_31

Frosini, L., Bardi, A., Manghi, P., & Pagano, P. (2018). An Aggregation Framework for Digital Humanities Infrastructures: The PARTHENOS Experience. *SCIRES-IT - SCIentific RESearch and Information Technology*, *8*(1). https://doi.org/10.2423/i22394303v8n1p33

Gladney, H. M. (1998, April). Data: Emerging trends and technologies. *D-Lib Magazine*. http://www.dlib.org/dlib/april98/04gladney.html

Graux, D., & Mahmoodi, S. (2021). A fully decentralized triplestore managed via the Ethereum blockchain. In *Further with Knowledge Graphs* (pp. 183-197). IOS Press.

Grosvenor, B. (2022, April 5). The British Museum's NFT project has sent its carbon footprint soaring. *The Art Newspaper*. https://www.theartnewspaper.com/2022/04/05/nfts-send-carbon-footprints-soaring-british-museum

Hakala, J. (2010). Persistent identifiers-an overview. *KIM Technology Watch Report*.

Haslhofer, B., & Isaac, A. (2011, July). Data.europeana.eu - The Europeana Linked Open Data Pilot. In *DCMI International Conference on Dublin Core and Metadata Applications*. The Hague, The Netherlands. Retrieved from http://eprints.cs.univie.ac.at/2919/

Hayward, A. (2023, January 4). NFT Sales in 2022 Nearly Matched the 2021 Boom, Despite Market Crash. *Decrypt*. https://decrypt.co/118438/2022-versus-2021-nft-sales

Henningsen, S., Florian, M., Rust, S., & Scheuermann, B. (2020, June). Mapping the interplanetary filesystem. In *2020 IFIP Networking Conference (Networking)* (pp. 289-297). IEEE.

Hern, A., & Sabbagh, D. (2021, June 8). Major internet outage 'shows infrastructure needs urgent fixing'. *The Guardian*. https://www.theguardian.com/technology/2021/jun/08/security-warning-error-cloud-websites-offline-outage

HM Treasury. (2022, April 4). *Government sets out plan to make UK a global cryptoasset technology hub*. GOV.UK. https://www.gov.uk/government/news/government-sets-out-plan-to-make-uk-a-global-cryptoasset-technology-hub

Humbel, M., Nyhan, J., Pearlman, N., Vlachidis, A., Hill, J. D., & Flinn, A. (in press). Socio-cultural challenges in collections digital infrastructures. *Journal of Documentation*. https://doi.org/10.1108/JD-12-2023-0263

IPFS. (2024). IPFS and Protocol Labs. Retrieved April 11, 2024, from https://docs.ipfs.tech/concepts/faq/#ipfs-and-protocol-labs

IPFS. (2024b). What is a CID?. Retrieved April 11, 2024, from https://docs.ipfs.tech/concepts/content-addressing/#what-is-a-cid

IPFS. (2024c). IPFS origins and a new P2P summer (2013 - 2017). Retrieved April 11, 2024, from https://docs.ipfs.tech/project/history/#ipfs-origins-and-a-new-p2p-summer-2013-2017

IPFS. (2024d). Performance. Retrieved April 11, 2024, from https://docs.ipfs.tech/concepts/ipfs-solves/#performance

Kaleido. (2023, December 5). IPFS: What You Need to Know. Retrieved April 11, 2024, from https://www.kaleido.io/blockchain-blog/what-is-ipfs

Kaur, G. (2023, July 15). What is Filecoin? The Leading Decentralized Storage Network. *CoinTelegraph Learn*. https://cointelegraph.com/learn/what-is-filecoin-the-leading-decentralized-storage-network

Kelly, M., Ivanovic, D., Rauch, C. B., Kunze, J., Grabus, S., Boone, J., Logan, P. M., & Greenberg, J. (2021). Archival Resource Keys for Collaborative Historical Ontology Publication. *Proceedings of the ICTeSSH 2021 Conference*. https://doi.org/10.21428/7a45813f.906ecfac

Khoury, D., Kfoury, E. F., Kassem, A., & Harb, H. (2018). Decentralized voting platform based on Ethereum blockchain. In *2018 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET)* (pp. 1-6). IEEE. https://doi.org/10.1109/IMCET.2018.8603050

Kotarski, R., Page, R., Padfield, J., & Mitchell, L. (2022). Persistent Identifiers as IRO Infrastructure. UK Research and Innovation. https://gtr.ukri.org/projects?ref=AH%2FT011092%2F1

Kunze, J. (2003). Towards Electronic Persistence Using ARK Identifiers. UC Office of the President. Retrieved from https://escholarship.org/uc/item/3bg2w3vs

Laddusaw, R., Sare, L., & Buckner, S. (2017). The Government Documents Digitization Initiative: Shepherding Resources from Shelf to Server. Texas Digital Library. http://hdl.handle.net/2249.1/82150

Manghi, P., Artini, M., Atzori, C., Baglioni, M., Bardi, A., La Bruzzo, S., De Bonis, M., Dimitropoulos, H., Foufoulas, I., & Iatropoulou, K., et al. (2017). OpenAIRE: Advancing open science. In *Proceedings of the Nineteenth International Conference on Grey Literature* (pp. 23-24). Rome, Italy.

Maguire, A. (2021, July 27). Internet outages are a timely reminder of why we need to decentralize the web. *Nasdaq*. https://www.nasdaq.com/articles/internet-outages-are-a-timely-reminder-of-why-we-need-to-decentralize-the-web-2021-07-27

Metilli D., Vlachidis A., MacDonald I., Lippolis A. S., Sadek J., Hughes A., Li J. & Nyhan J. (Forthcoming). Multivocality in the collection of everything: towards a Sloane Lab data model.

National Heritage Science Forum. (n.d.). RICHeS - Research Infrastructure for Conservation and Heritage Science. Retrieved April 1, 2024, from https://www.heritagescienceforum.org.uk/what-we-do/riches

Neumueller, A. (2023, April 26). New tool estimates environmental impact of blockchain networks. *Cambridge Judge Business School*. https://www.jbs.cam.ac.uk/2023/blockchain-sustainability-ethereum/

Niccolucci, F., Felicetti, A., & Hermon, S. (2022). Populating the Data Space for Cultural Heritage with Heritage Digital Twins. *Data, 7*(8), 105. https://doi.org/10.3390/data7080105

Nicholson, D., & Macgregor, G. (2003). NOF-Digi: Putting UK culture online. *OCLC Systems & Services: International Digital Library Perspectives, 19*(3), 96-99.

Nyhan, J., Vlachidis, A., Carine, M. A., Hill, J., Jansari, S., Pearlman, N., ... & Valeonti, F. (2023). The Sloane Lab: Looking back to build future shared collections [Project]. Zenodo. https://doi.org/10.5281/zenodo.7346795

ORCID. (n.d.). What are persistent identifiers (PIDs)? *ORCID Support*. Retrieved April 11, 2024, from https://support.orcid.org/hc/en-us/articles/360006971013-What-are-persistent-identifiers-PIDs

Poort, J., van der Noll, R., Ponds, R., Rougoor, W., & Weda, J. (2013). The value of Europeana: the welfare effects of better access to digital cultural heritage. (SEO-report; No. 2013-56). SEO economic research/Atlas voor gemeenten. http://www.seo.nl/uploads/media/2013-56_The_value_of_Europeana.pdf

Pernet, C., Horejsi, J., & Lu, L. (2023, March 16). IPFS: A New Data Frontier or a New Cybercriminal Hideout? Trend Micro. https://www.trendmicro.com/vinfo/au/security/news/cybercrime-and-digital-threats/ipfs-a-new-data-frontier-or-a-new-cybercriminal-hideout

Potts, J., & Rennie, E. (2019). Web3 and the creative industries: how blockchains are reshaping business models. In *A research agenda for creative industries* (pp. 93-111). Edward Elgar Publishing.

Protocol Labs. (2024). About. Retrieved April 11, 2024, from https://protocol.ai/about/

Parker, A. (2018, September 17). Distributed Web Gateway. *Cloudflare Blog*. Retrieved April 11, 2024, from https://blog.cloudflare.com/distributed-web-gateway/

Protocol Labs. (2017, July 19). Filecoin: A decentralized storage network [Whitepaper]. Retrieved from https://filecoin.io/filecoin.pdf

Purday, J. (2009). Think culture: Europeana.eu from concept to construction. *Bibliothek Forschung und Praxis*, *33*(2), 170-180. https://doi.org/10.1515/bfup.2009.018

Segundo, W., Matas, L., Nóbrega, T., Filho, J. E. S., & Mena-Chalco, J. (2023). *dARK: A decentralized blockchain implementation of ARK Persistent Identifiers* (No. 9516). EasyChair.

Shen, J., Li, Y., Zhou, Y., & Wang, X. (2019). Understanding I/O Performance of IPFS Storage: A Client's Perspective. In *2019 IEEE/ACM 27th International Symposium on Quality of Service (IWQoS)* (pp. 1-10). Phoenix, AZ, USA. https://doi.org/10.1145/3326285.3329052

Sicilia, M. A., Sánchez-Alonso, S., & García-Barriocanal, E. (2016). Sharing linked open data over peer-to-peer distributed file systems: the case of IPFS. In *Metadata and Semantics Research: 10th International Conference, MTSR 2016, Göttingen, Germany, November 22-25, 2016, Proceedings*(pp. 3-14). Springer International Publishing.

Sicilia, M. A., García-Barriocanal, E., Sánchez-Alonso, S., & Cuadrado, J. J. (2019). Decentralized Persistent Identifiers: a basic model for immutable handlers. *Procedia computer science*, *146*, 123-130.

Storage Market. (2024). The State of Storage. Retrieved April 11, 2024, from https://storage.market

Strassel, S., & Tracey, J. (2016, May). LORELEI Language Packs: Data, Tools, and Resources for Technology Development in Low Resource Languages. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 3273-3280). Portorož, Slovenia: European Language Resources Association (ELRA). Retrieved from https://aclanthology.org/L16-1521

Terras, M. (2012). Digitisation and digital resources in the humanities. In M. Terras, C. Warwick, & J. Nyhan (Eds.), *Digital Humanities in Practice* (pp. 47-70).

Tharani, K. (2021). Much more than a mere technology: A systematic review of Wikidata in libraries. *The Journal of Academic Librarianship, 47*(2), 102326. https://doi.org/10.1016/j.acalib.2021.102326

Towards a National Collection. (n.d.). Foundation Projects. Retrieved April 1, 2024, from https://www.nationalcollection.org.uk/Foundation-Projects

Towards a National Collection. (n.d.-b). COVID-19 Urgency Projects. Retrieved April 10, 2024, from https://www.nationalcollection.org.uk/Urgency

Towards a National Collection. (n.d.-c). Discovery Projects. Retrieved April 10, 2024, from https://www.nationalcollection.org.uk/Discovery_ProjectsValeonti, F., Bikakis, A., Terras, M., Speed, C., Hudson-Smith, A., & Chalkias, K. (2021). Crypto collectibles, museum funding and OpenGLAM: challenges, opportunities and the potential of Non-Fungible Tokens (NFTs). *Applied Sciences*, *11*(21), 9931.

Ugarte-Rojas, H., & Chullo-Llave, B. (2020). BLONDiE: blockchain ontology with dynamic extensibility. *CoRR, abs/2008.09518*.

UK Research and Innovation. (2024, March 22). Towards a national collection – opening UK heritage to the world. https://www.ukri.org/what-we-do/browse-our-areas-of-investment-and-support/towards-a-national-collection-opening-uk-heritage-to-the-world/

Valeonti, F. (2022, April 21). Are NFTs a real solution for museums? *Right Click Save*. Retrieved from https://www.rightclicksave.com/article/are-nfts-a-real-solution-for-museums

Van Zundert, J. (2012). If you build it, will we come? Large scale digital infrastructures as a dead end for digital humanities. *Historical Social Research/Historische Sozialforschung*, 165-186.

Verborgh, R., Vander Sande, M., Colpaert, P., Coppens, S., Mannens, E., & Van de Walle, R. (2014). Web-scale querying through linked data fragments. In *Proceedings of the LDOW*.

ViewBlock. (2024). Weave Size | Arweave Explorer. Retrieved April 11, 2024, from https://viewblock.io/arweave/stat/cumulativeWeaveSize

Yaga, D., Mell, P., Roby, N., & Scarfone, K. (2018). *Blockchain technology overview* (NIST IR 8202). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8202

Zhao, W., Rong, C., Wu, J., Sun, Z., & Sampalli, S. (2021). IEEE Access Special Section Editorial: Blockchain Technology: Principles and Applications. *IEEE Access, 9*, 110006-110010. https://doi.org/10.1109/ACCESS.2021.3101888