The Use of Blockchain for Digital Archives: a comparison between Ethereum and Hyperledger

[1]Angelica Lo Duca, [2]Clara Bacciu, [3]Andrea Marchetti

IIT CNR, Pisa, Italy
[1]angelica.loduca@iit.cnr.it
[2]clara.bacciu@iit.cnr.it
[3]andrea.marchetti@iit.cnr.it

Abstract

In recent years, blockchain technology is progressively spreading on a large scale in various research sectors, including Cultural Heritage. Different types of blockchain exist, which can be classified either according to the type of users that can access them, or based on the features they offer. This article describes a theoretical study in which two very different blockchains are compared: Ethereum and Hyperledger, in order to define which of the two is more suitable for storing tangible heritage contained in digital archives. After a brief description of the two technologies, a possible generic application scenario will be described in order to understand which of the two technologies best meets the requirements of the scenario. The comparison between the two blockchains will therefore be carried out on the basis of *general issues*, *architectural requirements* and *various considerations*. As a result of the comparison, it will emerge that Hyperledger Fabric is more suitable in the context of digital archives.

Negli ultimi anni la tecnologia blockchain si sta diffondendo sempre più su larga scala in diversi settori di ricerca, inclusi i Beni Culturali. Esistono diverse tipologie di blockchain, che possono essere classificate sia in base al tipo di utenti che possono accedervi, sia in base alle funzionalità che offrono. Questo articolo descrive uno studio teorico in cui si confrontano due blockchain molto diverse tra di loro: Ethereum e Hyperledger, al fine di definire quale delle due è maggiormente indicata per la memorizzazione di beni culturali tangibili contenuti in archivi digitali. Dopo una breve descrizione delle due tecnologie, verrà descritto un possibile scenario di applicazione abbastanza generico per poter capire quale delle due tecnologie meglio soddisfa i requisiti. Verrà quindi effettuato il confronto tra le due blockchain sulla base di *problematiche generali*, *requisiti architetturali* e *considerazioni varie*. Come risultato del confronto, emergerà che Hyperledger Fabric è più adatta nel contesto degli archivi digitali.

Introduction

Recently, the diffusion of applications based on blockchain technology [21], [25] has been increasing rapidly. The original focus of these technologies concerned cryptocurrencies (i.e., Bitcoin), but is shifting to finance and business in general, and is being extended progressively for a variety of applications in healthcare, government, Internet of Things, entity and assets management and eventually Cultural Heritage. In particular, a blockchain could be a good solution to store, protect and preserve over time data about tangible heritage, especially minor tangible heritage, i.e. artistically relevant artworks but not as famous as masterpieces. For example, in case of disasters (either natural or man made), the fact that the blockchain is a replicated registry can be exploited to retrieve information that in other circumstances would otherwise be lost forever. In addition, information contained in the blockchain cannot be erased or tampered with, so, in case of theft of the real artwork, related data will remain available and would be used to recognise the work if someone tries to sell it and to detect counterfeits. Using blockchain to store digital archives of artworks thus constitutes a promising field of application.

This paper is an extension of the work illustrated in [2]. In particular, it describes the challenges and requirements for storing a digital archive of artworks in a blockchain. In addition, a possible framework based on blockchain is illustrated. Then the paper describes a comparison between two blockchains, Ethereum[1] [24] and Hyperledger Fabric[2] [5], used as main framework for digital archives. A preliminary implementation of a framework for digital archives, based on Ethereum, has already been defined in [6], [3]. On the basis of the described framework, a selection of some comparison criteria is done, including *general issues* related to blockchains, *architectural requirements* applied to the specific architecture and *other considerations*. As a result of the comparison, we can say Hyperledger Fabric better fits for the proposed scenario because is more configurable. However, due to its popularity Ethereum still remains a good solution.

In addition to Ethereum and Hyperledger there are other implementations of the blockchain technology. Some of the most important are: Bitcoin [18], Corda,[3] Quorum.[4] Bitcoin was the first blockchain. Based on open source code, it implements a decentralized digital cryptocurrency where transactions are validated by miners through a rewarding process. Corda is an open source blockchain platform, designed mainly for business applications. Similarly, to Corda, Quorum, based on Ethereum with added control for permission and privacy, is a blockchain envisaged mainly for business applications. A complete comparison among the most important blockchains is done in [15]. The choice of which blockchain should be used to store digital archives depends mainly on two aspects: firstly, the blockchain should be *general*, i.e., not limited to financial applications. Secondly, it should be *popular*, i.e. technically mature

---

1   https://www.ethereum.org/

2   https://www.hyperledger.org/projects/fabric

3   https://www.corda.net/

4   https://www.jpmorgan.com/global/Quorum

and with community guaranteeing long-term sustainability.

This paper compares only Ethereum and Hyperledger Fabric, mainly because they represent the two main ways of implementing a blockchain: on the one hand Ethereum is the most representative example of all *permissionless* blockchains. On the other hand, Hyperledger represents *permissioned* blockchains. Both Ethereum and Hyperledger can be applied to specific scenarios through the use of smart contracts/chain codes.

Bitcoin is a digital currency, and programmability is very limited. Corda, instead, is a more recent blockchain, still not established as Ethereum and Hyperledger Fabric. Quorum is a specific implementation of Ethereum, thus some considerations done for Ethereum are valid also for Quorum.

Related Works

The problem of managing records through a blockchain has been largely investigated during the last few years. In her paper, Lemieux proposes a classification of blockchain applications [16], based on which information is stored in the blockchain: a) mirror type, b) digital record type, c) tokenized type.

Mirror type

In the mirror type, the blockchain serves as a mirror, which stores only records fingerprints. The complete information of a record is stored into an external repository and the blockchain is used only to verify records integrity. In [9] the authors describe a first implementation of a decentralized database for the storage of descriptive metadata related to digital records, based on the combination of the blockchain and IPFS technologies. In their paper Liang et. al. describe ProvChain [17], a system which guarantees data provenance in cloud environments. Vishwa et. al. [23] illustrate a blockchain-based framework, which guarantees copyright compliance of multimedia objects by means of smart contracts.

Digital record type

In the digital record type, the blockchain is used to store all the records in the form of smart contracts. In [4] the authors illustrate a distributed and tamper-proof framework for media. Each media is represented by a watermark, which is firstly compressed and then stored into a blockchain. Approved modifications to media are stored in the blockchain thus preventing tampering. In [8] the authors describe Archain, a blockchain-based archive system, which stores small sized records. Multiple roles are defined in the system, thus allowing records creation, approval and removal.

Tokenized type

In the tokenized type, records are stored in the blockchain and they are linked to a cryptocurrency. Adding, updating or removing a record has a cost. This constitutes an innovative case, where the literature is not consolidated yet. An example of this type of blockchain is represented by the Ubitquity Project,[5] which records land transactions on behalf of companies and government agencies.

Background

The concept of Digital Archive

A digital archive is a repository of digital records that need long-term or even permanent preservation for their cultural, historical, or evidentiary value. In digital archives, a record can be anything holding a piece of information in the form of digital object, such as texts, images, pictures, videos and audios. This paper focuses on digital archives which contain collections of minor artworks. Minor artworks are artistically relevant works but not as well-known as famous masterpieces, or belonging to the so-called minor arts, such as books and manuscripts, pottery, lacquerware, furniture, jewellery, or textiles. Examples of minor tangible heritage could be those kept in some small libraries or countryside churches, or even in private households.

The creation, management and sustainability of a digital archive is not an easy task, because there is a series of issues that must be taken into consideration [1], [13], [26], [11]. The InterPARES (International Research on Permanent Authentic Records in Electronic Systems) series of projects [22] focused on creating policies and guidelines for making and maintaining digital records, including authenticity requirements for record systems and long-term preservation of digital records.

A digital archive is subject to obsolescence, in the sense that the hardware supports on which it is stored change over time (from the floppy disk to the Internet cloud). Thus a digital archive needs *long-term preservation*, i.e. digital artwork should remain accessible for a long period of time depending on legal, regulatory, operational, and historical requirements. Secondly, every artwork of the digital archive must be associated with different metadata (descriptive, structural, administrative), which should be maintained up-to-date by authorized accounted persons. This means that on the one hand that all the operations about the digital archive should be documented in an open and verifiable manner (*transparency*). On the other hand, artworks should be protected against forgery and identified correctly in case of loss and subsequent discovery (*anti-counterfeiting*). Thirdly, records of the digital archive are stored in different media formats, each defined by its own software and hardware. A digital archive should guarantee the availability of all the formats, i.e. artworks should be efficiently and accurately retrieved. Finally there are also other aspects that must be considered, such as corruption and loss of information, which need *protection*, *integrity* and *traceability* of artworks.

---

5   http://ubitquity.io/brazil_ubitquity_llc_pilot.html

Integrity makes sure that the digital description of the artwork is not subject to unauthorized changes. Protection permits to protect the digital description of the artwork in case of natural disasters and/or attacks (it is obviously impossible to protect the real work only with IT tools). Traceability permits to trace all movements of individual artworks.

Another relevant issue concerns how difficult it can be to find and access repositories due to the inconsistent description practices among different archives. The ISAD(G) (General International Standard Archival Description) [7] is a standard that addresses this issue and gives guidelines, to be used in conjunction with existing national standards, for the preparation of archive descriptions that are effective in presenting the content of archival material, so that it is easily identifiable and accessible. This description creates a hierarchy of metadata related to the entire archive, as opposed to those related to each record.

An overview of blockchain technology

A blockchain is a particular implementation of a Distributed Ledger (DL). A DL is essentially a database, which is shared among different nodes of a network. In practice, all the nodes of the network share the same copy of the database and any change made on a node, is replicated to all the other nodes in few minutes and, in some cases, even in few seconds. A DL can be public (as opposite of private) if any node can read the content, and permissionless (as opposed of permissioned) if any node can write content (Table 1).

The protocol for the first functioning blockchain was introduced in 2008 to support the digital cash Bitcoin, and implements the ledger as a chain of blocks. Each block contains data, a timestamp and a cryptographic hash of the previous block. This way the integrity of the information stored in the blockchain is protected through a security system based on cryptography.

With respect to a standard database, a blockchain is an append-only register. This means that information can only be added to the database, but it cannot be removed. Modifications to the stored data can be done by re-uploading a new version of the data. A distributed consensus algorithm is used to decide which updates to the ledger are to be considered valid. New participants (nodes) can start collaborating to the maintenance of the repository by following this algorithm. There is no need of a central authority or trust between nodes; the consensus algorithm and cryptography grant the correctness of data even in the presence of some malicious nodes.

Each block is made tamper-resistant by adding in its header a cryptographic signature of the data it contains (usually a hash of the content), as well as a link to the previous block of the chain (the cryptographic hash of the block). This way each block is dependent on the content of all the previous blocks, making it impossible to modify the data contained in old blocks without rewriting the new ones.

Initially designed for financial transactions, blockchain technology can be used to record anything of value. Even executable code can be stored in the blockchain, the so-called smart

contracts. A smart contract is not necessarily the transposition of a real contract, it is just code that is executed by all the nodes of the blockchain network, and the result of the computation is stored after a consensus is reached. A transaction carrying the payload of the contract is first broadcast to the network. Its result is the deployment of the payload as code linked by its public address. Any new transaction can then refer to this address to trigger the execution of the functions inside the contract.

The big advantage of a blockchain is that it is an immutable, distributed, always available, secure and publicly accessible repository of data. The main issues with blockchain implementation of distributed ledgers are scalability and efficiency: often, consensus algorithms that are used to grant consistency are expensive in terms of time and resources. In some cases, a certain level of trust among participants can be present, thus simpler consensus algorithms can be used.

Key technical choices of blockchain technology include:

1) permission design, i.e., whether permission is needed to access the blockchain;

2) choice of consensus algorithm, i.e., how a new block is added to the blockchain;

3) whether or not to use smart contracts, i.e., whether to use the blockchain as a virtual machine where programs representing business processes are run;

4) whether or not to use a cryptocurrency, i.e., whether the consensus algorithm and smart contract operations depend on an artificial currency or not.

Those technical choices often result from the governance model that has been chosen for the ecosystem of participants.

	SOME	ALL
CAN READ	PRIVATE	PUBLIC
CAN WRITE	PERMISSIONED	PERMISSIONLESS

Table 1: Types of blockchains according who can access what.

Ethereum

Ethereum is a public open-source blockchain platform that has the capability of running so-called decentralised applications (dApps). At the moment, the consensus algorithm is based on Proof of Work. Mining nodes generate a cryptocurrency named Ether that is used to pay for transactions. The key characteristic of Ethereum is that it is a programmable blockchain, because it provides a Virtual Machine (EVM) that can execute user generated scripts (smart

contracts) using the network of nodes. Smart contracts are usually written in Solidity[6] language (but there are some alternatives), are compiled to EVM bytecode, and are deployed to the blockchain for execution. Contract computation consumes gas, which is paid spending Ether. Smart contracts are the foundation of dApps. Diagram in Figure 1 shows the simplified architecture of dApps: there is no central server to which Web every browser has to connect, but instead each one has its own instance of the application. Ethereum functions both as storage for data and code, and as the machine that executes the code.
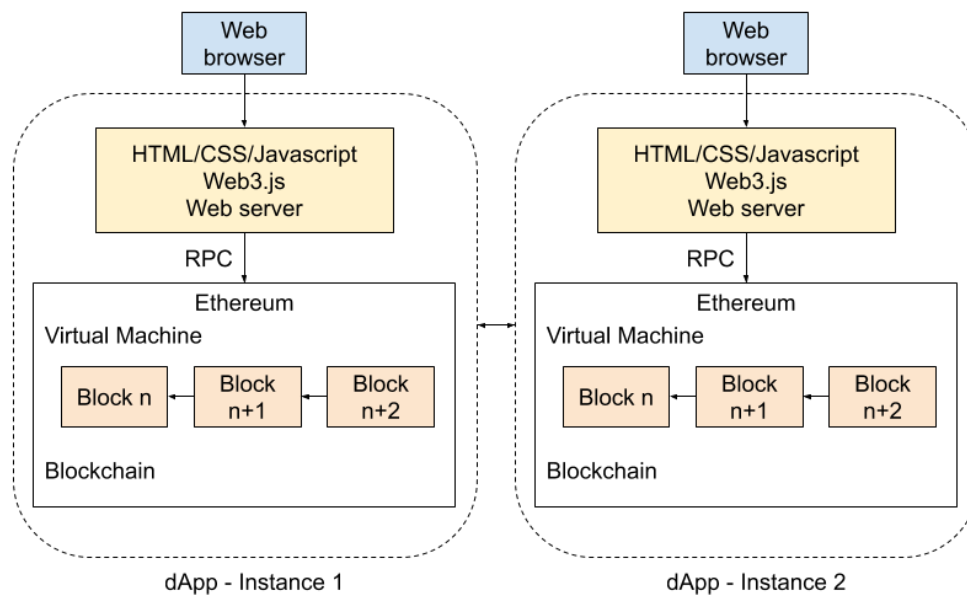

Figure 1: The architecture diagram of an Ethereum Dapp.

The Ethereum Network

The Ethereum network is a public distributed network with two types of nodes: *full nodes* and *lightweight nodes*. Full nodes which contain the whole blockchain, i.e. all the validated transactions. Some full nodes, called *miners*, are also responsible for transaction validation. Miners can also be grouped in *pools*. *Lightweight nodes* contain a subset of the blockchain and rely on full nodes for missing information. Examples of lightweight nodes are *e-wallets*, i.e. electronic devices or apps which permit to do transactions.

---

6   https://solidity.readthedocs.io

Ether and Gas

As already said, Ethereum is cryptocurrency-based blockchain where the cryptocurrency used is called Ether (ETH). The price of 1 ETH is 182.31 $ (updated on October 31, 2019). Together with Ether there is also Gas, which is used to pay computational resources in the network (*gas fee*). The current value of Gas is called *gas price*. Every smart contract has associated a *gas limit*, which is the maximum amount of gas which it can consume.

Hyperledger

Hyperledger is an open source effort aimed at advancing cross-industry blockchain technologies. Hyperledger focuses on developing different blockchain frameworks and modules to support global enterprise solutions. Hyperledger blockchains are generally permissioned blockchains, which means that the parties that want to join the network must be authenticated and authorized. The focus of Hyperledger is to provide a transparent and collaborative approach to blockchain development. Within Hyperledger, there are eight different technology code projects, which define a common set of development principles: five distributed ledger frameworks and three support modules. The Hyperledger frameworks include:

An append-only distributed ledger

A consensus algorithm for agreeing to changes in the ledger

Privacy of transactions through permissioned access

Smart contracts to process transaction requests.

In this paper only Hyperledger Fabric is described, because it is the most widespread. The Hyperledger Fabric blockchain is a distributed system consisting of many nodes that communicate with each other. Figure 2shows the Hyperledger Fabric Model.
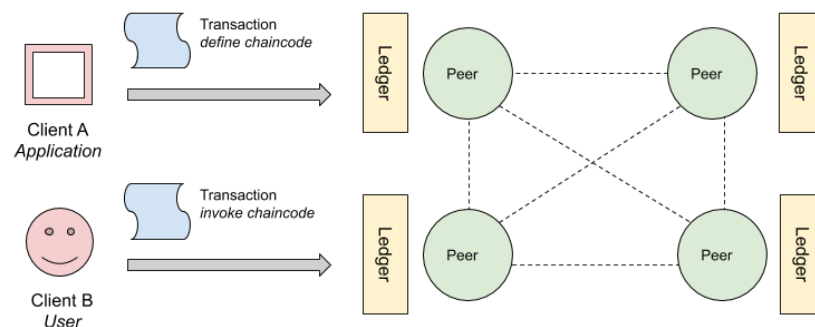

Figure 2: The Hyperledger Fabric Model. Client A defines a chaincode (contract) through a transaction. Once the transaction approved, client B can invoke methods contained in the chaincode through another transaction.

Chaincodes and channels

The blockchain runs programs called chaincodes, holds state and ledger data, and executes transactions. Chaincodes correspond to the Ethereum smart contracts. Each chaincode can be invoked through one or more operations, called transactions. Transactions have to be *endorsed* and only endorsed transactions may be committed and have an effect on the state of the ledger.

The most peculiar aspect of Hyperledger is the possibility to define *channels*, which are data partitioning mechanisms that allow transaction visibility for only some defined users of the blockchain. Each channel is an independent chain of transaction blocks containing only transactions for that particular channel.

The ledger contains the current world state of the network and a chain of transaction invocations. The world state reflects the current data about all the assets in the network. Ledger provides a verifiable history of all successful state changes (valid transactions) and unsuccessful attempts to change state (invalid transactions), occurring during the operation of the system.

Roles and transactions

In Hyperledger two roles can be defined: *clients* and *validators*. Clients are applications that act on behalf of a person to propose transactions on the network. Validators maintain the state of the network and a copy of the ledger.

Unlike Ethereum, in Hyperledger Fabric there is no mining of blocks. In order to verify a transaction, each transaction is sent to one trusted validator, which broadcasts it to all the other validators of the network. All the validators reach consensus (using a specific algorithm) on the order to follow to execute all the transactions. Then each validator runs the transactions on its own, following the established order and builds a block with all the executed transactions. Since the execution of transactions is deterministic, all the validators build exactly the same block. Finally, the validators asynchronously notify the client application of the success or failure of the transaction. Clients are notified by each validator.

The model of blockchain for digital archives

The use of blockchain for digital archives guarantees a mechanism to access, manage and protect cultural heritage on a daily basis and at times of disasters (due for example to climate change or man-made). The blockchain-based framework should be designed both for minor tangible heritage and major tangible and intangible heritage.

Thanks to the append-only-register property of the blockchain, the framework provides a layered protection and conservation means for cultural heritage. The framework exploits also some specific advantages of blockchain (integrity, transparency and authenticity of records) to allow the secure storage of minor tangible heritage contained in digital archives. The framework integrates also technologies for a distributed record storage, such as the InterPlanetary File

System[7] [14], in order to guarantee the digital preservation and transmission of tangible and intangible heritage from generation to generation. The use of these storage technologies permits the development of a sustainable protection and enhancement of values as well as the long-term management of cultural heritage at risk.

Thanks to the benefits of blockchain and the distributed technologies for record storage, the framework should improve sustainable access to digital heritage by contributing to the resilience of our societies in terms of:

helping users to preserve the memory of cultural heritage in case of destruction of the physical artwork, due to natural disasters or man-made disasters,

facilitating the restoration and/or the reconstruction of damaged heritage, thanks to the information contained in the ledger,

preventing malicious changes to the ledger,

registering temporary movements of movable heritage, for example for exhibits.

Requirements

The section *The concept of Digital Archive* describes the *general requirements* of a digital archive (long-term preservation, transparency, anti-counterfeiting, protection, integrity and traceability). The use of blockchain for digital archives should guarantee also the following *architectural requirements*:

interoperability: this aspect should guarantee that the blockchain can easily interoperate with external modules, such as web interfaces and external storage (i.e. IPFS);

customizable infrastructure: the system should guarantee that the underlying infrastructure is customizable, e.g. the number of nodes and costs can be decided independently;

roles: the blockchain should define different users roles, according to what specified in the previous section.

queries: this aspect refers to the ability to search data in the blockchain, e.g. search an artwork by title or author.

In addition to these requirements, the following parameters that affect performance and scalability should be taken into account [19]:

block frequency: inversely proportional to the time between two succeeding blocks. It is affected by mining difficulty;

block size: the number of transactions that fit in a block;

network size: the number of nodes of the network. Increasing the number of nodes in the network does not always improve performance. In fact, communication and consensus costs may increase.

---

7   https://ipfs.io/

throughput: the number of transactions per second;

latency: the time elapsed between the submission of a transaction and its validation;

finality: the property that once a transaction is completed, there is no way to alter it.

Architecture

Figure 4 describes the architecture of the blockchain-based framework for digital archives. Starting from the bottom of the figure, there is a Data Lake where all the descriptions of artworks are stored. The Data Lake is a distributed storage. Artworks in Data Lake can be accessed through indexes contained in the blockchain. The blockchain contains also other basic information related to artworks descriptions, as well as a track of all the operations done on each artwork. This means that an external audit of the framework can always verify the status of an artwork and determine if something is wrong. The blockchain constitutes the backend of the framework, together with the Cache. The Cache service stores basic information about artworks, such as author name and description, in order to make users queries faster. In fact, natively, a blockchain is not suitable for fast queries such those required by a web search engine.

The frontend of the framework is composed of the Authentication Service, which manages users access to the system, and three interfaces, one for each type of user: Search, Publisher, and Admin Interface. Users of the framework should play one of the following roles: *generic user*, *publisher, verifier*. A generic user can search for an approved artwork in the system. A publisher user can publish or update an artwork in the system. When a new artwork is published, its status is set to pending. This means that the artwork is not approved yet thus cannot be accessed by third parties neither can be updated by its author. A verifier is an expert in the field to which the artwork belongs, and can vote for the approval of the artwork description. This mechanism constitutes an algorithm for compliance with the principle of reliability of a record. Complex strategies can be defined to establish how an artwork should be approved.
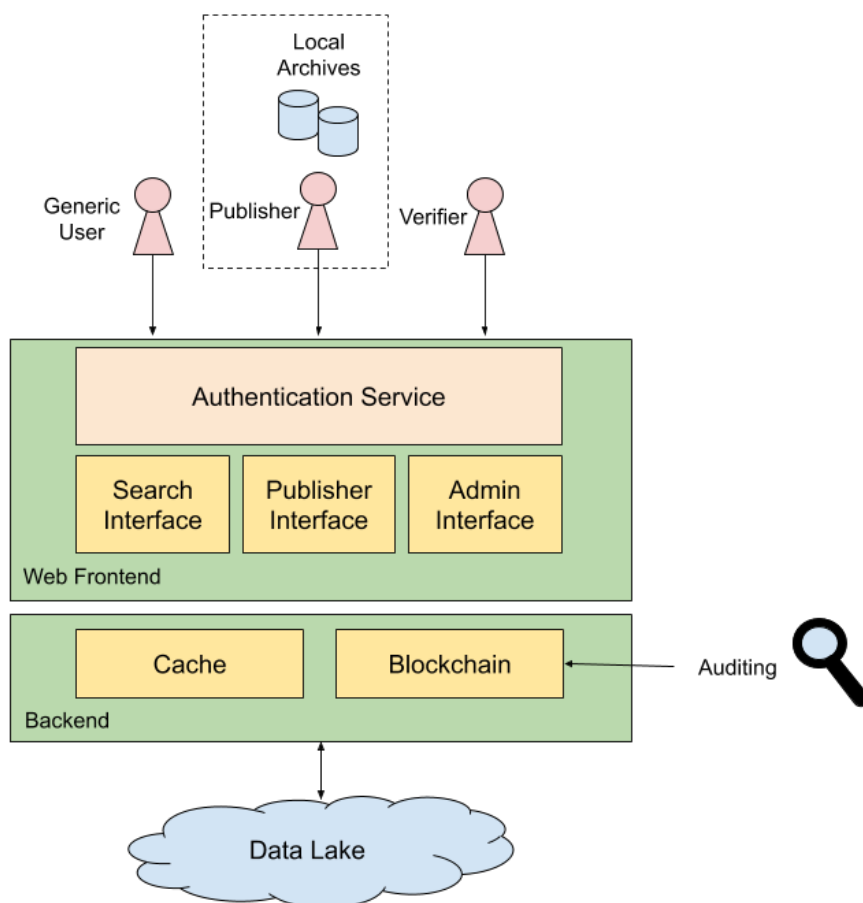
Figure 3: The architecture of the blockchain-based framework.

Due to its intrinsic nature, the blockchain already satisfies the general requirements of a digital archive, but long-term preservation, which is guaranteed through the Data Lake. Protection would be achieved through the fact that the blockchain is replicated on different nodes. Anti-counterfeiting would be guaranteed by associating each work to a sort of digital identity card, containing all the information related to the work (including physical information). Finally, integrity and traceability would be intrinsically guaranteed by the immutability and timestamping properties of the blockchain. In fact, blockchain security assumptions guarantee that if at a certain time a piece of information has been added to a block that reached consensus, it will be impossible to alter that information without altering all the following blocks.

Both Ethereum and Hyperledger Fabric satisfy the architectural requirements. However, depending on the type of blockchain, there are the following additional *issues* that should be taken into account:

costs: whether or not every transaction has a fee;

popularity: how the blockchain is known, i.e. there is a supporting community and there are skilled programmers able to implement contracts;

consensus: the distributed process which establishes the validation of transactions.

Discussion

Table 2 illustrates issues associated with the proposed architecture for digital archives and how the two blockchains address them. The fourth column of the table shows which blockchain fits better the requirements for digital archives. Regarding costs, Hyperledger Fabric fits better to the proposed architecture, because the network can be configured without costs on transactions. This means that all the categories of users can access the blockchain freely. However, if a business model were defined in the architecture, e.g. pay as you publish/access resources, also Ethereum could be suitable for digital archives. Anyway, a private network can be always set up on Ethereum, with a gas price set to zero, thus satisfying the model of the proposed architecture.

When dealing with popularity, Ethereum is more popular and well-known than Hyperledger Fabric. This means that a technical problem in the implementation of the described architecture could find a greater support by the Ethereum community that the Hyperledger one.

Regarding consensus, Ethereum bases it on Proof of Work, while Hyperledger Fabric implements a permissioned voting-based consensus that implies a level of trust among participants ad requires messages to be exchanged between nodes. In this case, Ethereum seems to be more fit because of the lack of need for trust among participants.

Summarizing, from the point of view of issues, Ethereum seems to behave better than Hyperledger Fabric, but for costs.

Issue	Ethereum	Hyperledger Fabric	Preferred blockchain
Costs	Every transaction has a cost dependent on gas price (current is about 4 gwei). In a private blockchain, the gas price can be configured.	Costs can be established when configuring the network	Hyperledger Fabric
Popularity	A well-established and wide community exist. Many programmers are able to write smart contracts	A developing niche community exists.	Ethereum
Consensus	Based on Proof of Work. The larger the network, the more reliable the consensus	Consensus algorithm requiring a level of trust and message overhead. The larger the network, more time it takes to reach consensus.	Ethereum

Table 2: Considerations about issues in the two blockchains and which one is more suitable for digital archives.

Table 3 describes the architectural requirements and how the two blockchains satisfy them. Like in the previous table, the fourth column specifies the preferred blockchain for a given requirement. Firstly, referring to the customizable architecture, Hyperledger Fabric is to prefer, because it permits to define who can access the network (permissioned blockchain). However, Ethereum can be set up as a private blockchain, in the sense that there is a single organization which manages it, and contracts handling user permissions can be implemented. Secondly, looking at interoperability with external storage, Hyperledger Fabric is better than Ethereum, because it has a native storage (data lake) and there is no need to configure external libraries to access it. Anyway, if a programmer has good skills with Ethereum, the configuration of external libraries to access external storage should not be difficult. The same analysis can be done for interoperability with Web Interfaces. Thirdly, regarding roles, Hyperledger Fabric is more configurable than Ethereum because of its native support of roles. Through channels, Hyperledger Fabric can define also more complex access policies. When defining roles, Ethereum has an overhead in terms of smart contracts thus is not indicated for the proposed

architecture. Finally, queries are not supported neither by Ethereum nor by Hyperledger and this aspect constitutes a limit of all blockchains. Thus, an additional mechanism based on caching is defined in the architecture, in order to speed up data searches. Summarizing the comparison about the architectural requirements, Hyperledger Fabric is the best solution.

Requirement	Ethereum	Hyperledger Fabric	Preferred blockchain
Customizable infrastructure	Ethereum is natively a public network, however a private blockchain can be set up	Native support of permissioned blockchain	Hyperledger Fabric
Interoperability with external storage (e.g. IPFS)	External libraries exist to support IPFS (e.g. Infura)[8]	No support to external storage because nodes in hyperledger fabric have already a local storage	Hyperledger Fabric, but Ethereum is a good alternative
Interoperability with Web Interfaces	External libraries exist (e.g. web3.js[9] and Drizzle)[10]	Native support of interoperability with web interfaces (in Angular JS)[11]	Hyperledger Fabric, but Ethereum is a good alternative
Roles	A smart contract must be defined to manage roles	Native support of roles through the definition of policies	Hyperledger Fabric
Queries	No native support	No native support	-

Table 3: Considerations about architectural requirements in the two blockchains and which one is more suitable for digital archives.

A direct comparison between Ethereum private and Hyperledger in terms of performance is

---

8   https://infura.io/

9   https://web3js.readthedocs.io/en/v1.2.2/

10   https://www.trufflesuite.com/docs/drizzle/quickstart

11   https://angularjs.org/

difficult, due to the fact that are both highly configurable.

In general, how the protocol and the network are configured (starting from the choice of a consensus algorithm) has a big impact on performance. An increase in mining difficulty leads to a block frequency decrease, throughput decrease and latency rise. Block size and block frequency are to be balanced (especially with PoW): incrementing block size improves performance only if the block period is large enough for nodes to be able to create, sign, propagate, execute transactions and reach consensus. Adding nodes to the network increases computation capability, but if block frequency is high and block size is large, some nodes may not have the resources to propagate information on time and keep in sync. Therefore, scaling is limited by the design of the blockchain platform.

A study about performance and scalability of Ethereum private networks can be found in [19], while performance metrics about Hyperledger are studied in [10].

As for the Ethereum public platform, average block frequency is 10-20 seconds, and average block size is 20-30 KB. Troughput is about 15 transactions per second, with a latency of about 6 minutes. Current network size is 2.717.215 nodes.[12]

Summarizing, although Ethereum is more popular than Hyperledger Fabric, Hyperledger Fabric seems more suitable to store digital archives, because it is highly configurable and permits to define roles natively, without additional overhead. However, as the first implementation of the proposed architecture, Ethereum is more indicated because of its simplicity and popularity [6], [3].

Conclusions and future work

This paper has presented challenges and requirements of storing digital archives through blockchain. In addition, a possible architecture based on blockchain has been illustrated as well as a preliminary comparison between Ethereum and Hyperledger Fabric as underlying blockchains in a framework for storing, protecting and preserving digital archives. The paper has compared the two blockchains at three levels: *natively issues* of the blockchain, *architectural requirements* and *general considerations*. As a result, Hyperledger Fabric is more suitable to store digital archives because of its high configurability.

We are aware that this study is preliminary, but we believe that the effort to define a possible architecture of the framework, as well to select and analyse which parameters should be considered when comparing two or more blockchains for digital archives storage is useful in this field.

As already said in the introduction, an implemented use case of this architecture can be found in [24], which exploits Ethereum as underlying blockchain. The further step should be the implementation of the framework in Hyperledger Fabric and then a comparison of the two implemented use cases.

---

12   https://etherscan.io/nodetracker/nodes

References

[1] ARMA. 2017. "International. Generally Accepted Recordkeeping Principles".
https://rim.ucsc.edu/management/images/ThePrinciplesMaturityModel.pdf

[2] Bacciu, Clara, Angelica Lo Duca and Andrea Marchetti. 2019. "The Use of
Blockchain for Digital Archives: Challenges and Perspectives". AIUCD Annual
Conference – Pedagogy, Teaching and Research in the Age of Digital Humanities, 78-
81.

[3] Basile, Mariano, Gianluca Dini, Andrea Marchetti, Clara Bacciu and Angelica Lo
Duca. 2019. "A blockchain-based support to safeguarding the Cultural Heritage".
EVA Proceedings of the Electronic Imaging & the Visual Arts, edited by V.
Cappellini,64-73. Firenze: FUP.

[4] Bhowmik, Deepayan, and Tian Feng. 2017. "The multimedia blockchain: A
distributed and tamper-proof media transaction framework." In DSP Proceedings of
the 22nd International Conference on Digital Signal Processing, 1-5. DOI:
10.1109/ICDSP.2017.8096051

[5] Cachin, Christian. 2016. "Architecture of the hyperledger blockchain fabric." In
Proceedings of the Workshop on Distributed Cryptocurrencies and Consensus
Ledgers. https://www.zurich.ibm.com/dccl/papers/cachin_dccl.pdf

[6] Clara Bacciu, Angelica Lo Duca, and Andrea Marchetti. 2019. "A Blockchain-based
Application to Protect Minor Artworks." In Proceedings of the 15th International
Conference on Web Information Systems and Technologies, edited by A. Bozzon, F.
Dominguez Mayo and J. Filipe, 319-325. Setúbal: Scitepress.
DOI:10.5220/0008347903190325

[7] Duranti, L., and R. Preston. 2008. International research on permanent authentic
records in electronic systems (InterPARES) 2: Experiential, interactive and dynamic
records. Padova: CLEUP.

[8] Galiev, Albert, Shamil Ishmukhametov, Rustam Latypov, Nikolai Prokopyev, Evgeni
Stolov and Ilya Vlasov. 2018. "Archain: a novel blockchain based archival system." In
WorldS4 Proceedings of the Second World Conference on Smart Trends in Systems,
Security and Sustainability, 84-89. DOI: 10.1109/WorldS4.2018.8611607

[9] García-Barriocanal, Elena, Salvador Sánchez-Alonso and Miguel-Angel Sicilia. 2017.
"Deploying metadata on blockchain technologies." In Research Conference on
Metadata and Semantics Research, 38-49. https://doi.org/10.1007/978-3-319-70863-
8_4

[10] https://doi.org/10.1007/978-3-030-30429-4_

[11] Hyperledger Performance and Scale Working Group. 2018. "Hyperledger Blockchain
Performance Metrics" Whitepaper. https://www.hyperledger.org/wp-

content/uploads/2018/10/HL_Whitepaper_Metrics_PDF_V1.01.pdf

[12] International Council on Archives / Conseil international des archives. 2000. ISAD (G): General International Standard Archival Description, Second Edition, Adopted by the Committee on Descriptive Standards, Stockholm, Sweden, 19-22 September 1999. Ottawa.

[13] ISO. 2016. "ISO 15489-1/2: 2016- Information and documentation - Records management". https://www.iso.org/standard/62542.html

[14] Juan Benet. 2014. "Ipfs-content addressed, versioned, p2p file system." arXiv preprint. arXiv:1407.3561.

[15] Kuny, T. 1998. "The digital dark ages? Challenges in the preservation of electronic information." International preservation news 17:8-13.

[16] Lemieux, Victoria. 2017. "A typology of blockchain record- keeping solutions and some reflections on their implications for the future of archival preservation." In Big Data 2017. Proceedings of the IEEE International Conference on Big Data, 2271–2278. DOI: 10.1109/BigData.2017.8258180

[17] Liang, Xueping, Sachin Shetty, Deepak Tosh, Charles Kamhoua, Kevin Kwiat, Laurent Njilla. 2017. "Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability." In CCGRID 2017 Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, 468-477. DOI: 10.1109/CCGRID.2017.8

[18] Nakamoto, Satoshi. 2008. "Bitcoin: A peer-to-peer electronic cash system". Website. https://bitcoin.org/bitcoin.pdf

[19] Schäffer, Markus, Monica Di Angelo and Gernot Salzer. 2019. "Performance and Scalability of Private Ethereum Blockchains." In BPM 2019 Proceedings of the Business Process Management: Blockchain and Central and Eastern Europe Forum. DOI

[20] Schwartz, David, Noah Youngs and Arthur Britto. 2014. "The ripple protocol consensus algorithm". Ripple Labs Inc WhitePaper. https://ripple.com/files/ripple_consensus_whitepaper.pdf

[21] Swan, Melanie. 2015. "Blockchain: Blueprint for a new economy". O'Reilly Media.

[22] Tsung-Ting Kuo, Hugo Zavaleta Rojas, Lucila Ohno-Machado. 2019. "Comparison of blockchain platforms: a systematic review and healthcare examples." Journal of the American Medical Informatics Association 26, no. 5:462–478.

[23] Vishwa, Alka, and Farookh Khadeer Hussain. 2018. "A blockchain based approach for multimedia privacy protection and provenance." In SSCI 2018 Proceedings of the IEEE Symposium Series on Computational Intelligence, 1941-1945. DOI:10.1109/ssci.2018.8628636

[24] Wood, Gavin. 2014. "Ethereum: A secure decentralised generalised transaction ledger". Ethereum Project Yellow Paper. https://gavwood.com/paper.pdf

[25] Zeilinger, Martin. 2018. "Digital art as 'monetised graphics': Enforcing intellectual property on the blockchain." Philosophy & Technology 31, no. 1: 15-41. DOI: 10.1007/s13347-016-0243-1

[26] Zheng, Zibin, Shaoan Xie, Hong-Ning Dai, Xiangping Chen and Huaimin Wang. 2018. "Blockchain challenges and opportunities: A survey." In International Journal of Web and Grid Services, 14/4, 352-375.

Last access URLs: 28th October 2019.