The Use of Blockchain for Digital Archives: Challenges and Perspectives

ABSTRACT

Over the last few years, the exploitation of blockchain technologies is increasing across different sectors, such as health and economy. A field that still remains little explored is that of digital archives owned by public administrations and other authoritative organizations. This paper investigates challenges and perspectives about the use of blockchain technologies for digital archives.

The paper describes also a possible application scenario, which employs blockchain technology for the registration of minor artworks. The described platform deals with problems of artwork counterfeiting, loss and subsequent discovery, natural disasters and traceability.

KEYWORDS

Blockchain; Digital Archives; Digital Objects.

1. INTRODUCTION

Recently, the diffusion of applications based on blockchain technology [7] 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. A field which still seems to remain little explored is Cultural Heritage.

We would like to investigate on the possible challenges and benefits derived by the application of blockchain technologies to this field, and, more in general, to digital archives.

The application of blockchain to digital archives and in general to Cultural Heritage is very recent thus literature about this topic is still little. The ARCHANGEL project[1] will implement trusted archives of digital public records on a distributed ledger. The project has started last year and it will end by the end of 2018.

Memory Matrix[2] is an ongoing project aiming at inserting information about Cultural Heritage in the Bitcoin network.

The Kapu project[3] is another ongoing activity, which tries to create a blockchain for archeology.

The potential benefits introduced by the use of blockchain for digital archives are essentially the following:

a) replication, which prevents loss of information; b) documented provenance; c) immutability and resistance to unauthorized changes.

2. AN OVERVIEW OF BLOCKCHAIN

A blockchain is a particular implementation of a Distributed Ledger (DL) [9]. 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.

The protocol for the first functioning blockchain was introduced in 2008 to support the digital cash Bitcoin [5], 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

[1] http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=EP/P03151X/1

[2] http://www.memorymatrix.org

[3] https://www.kapu.one

authority or trust between nodes; the consensus algorithm and cryptography grant the correctness of data even in presence of some malicious nodes.
Among the most important blockchain protocols are: the above mentioned Bitcoin, Ethereum [8], Hyperledger [3], Ripple Transaction Protocol (RTXP) [3].

3. THE CONCEPT OF DIGITAL ARCHIVE

Archives contain records or collections of 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.
The creation, management and use of a digital archive is not an easy task. ARMA International[4]'s Generally Accepted Recordkeeping Principles [1] define a global standard that identifies the criticalities and a high-level framework of good practices for information governance. They are a common set of principles that describe the conditions under which business records and related information should be maintained. They are:
Accountability: there should be a person that is responsible and accountable for all the process;
Transparency: all the information should be documented in an open and verifiable manner;
Integrity: the information assets should be as authentic and reliable as possible;
Protection: no unauthorized parties should be able to access private information;
Compliance: laws and policies should be kept into consideration;
Availability: information should be efficiently and accurately retrieved;
Retention: information should remain accessible for a period of time depending on legal, regulatory, fiscal, operational, and historical requirements;
Disposition: it should be possible to erase all the information that is no longer needed.

The ISO Standard 15489-1:2016 Information and documentation — Records management [4] defines concepts and principles for the creation, acquisition and management of records. Section 7.2 *Characteristics of a record* lists the following:
Authenticity: records must be created and maintained in such a way that creators are authorized and identified, and that records are protected against unauthorized addition, deletion, alteration, use and concealment
Reliability: the content of a record should be accurate, and its creator should be worth of trust
Integrity: a record should be complete and protected against unauthorized alteration. Every alteration should be documented and traceable
Usability: a useable record is one that can be located, retrieved, presented and interpreted.

4. CHALLENGES AND PERSPECTIVES

Even if those guidelines have been defined having in mind mainly the management of archives in business activities, some of them are very generic and can be referred to any digital archive. The most relevant features are reliability, authenticity, and availability over time. Reliability has a lot to do with the trustworthiness of the creator of a record, and their competence to capture the truth, so it is necessary to define an effective method of authentication that avoids problems of plagiarism. Authenticity regards the identity of a record, and its integrity. Finally, availability over time means providing long-term access to the information. Long-term refers to a period of time that is long enough to survive to changing technologies, such as new media and data formats.
Ensuring a correct access to digital archives includes the following challenges:

Digital Preservation
Digital preservation guarantees that digital objects are available and usable over the time [2]. A digital archive should take care of storage media instability and deterioration, which could lead to data loss, and technology obsolescence and incompatibility, which may happen both at the hardware and software level.

Decentralization
A digital archive is often owned by a centralized organization, such as a government or a library, which stores all digital objects in a local database and then publishes them on the Web. This means that the archive leaves as long as the centralized organization owning it. Decentralization should guarantee that the ownership of a digital archive is not associated to any

[4] ARMA International is a not-for-profit professional association and a global authority on governing information as a strategic asset

particular organization. Digital objects should belong to the human heritage and they should be accessible and available everywhere and every time, despite the organization hosting the corresponding physical copies.

Security Issues

Security should guarantee the integrity of digital objects contained in the archive, i.e. digital objects should not be modified by not authorized entities. In addition, a secure digital archive should prevent plagiarism, i.e. it should guarantee that only authorized digital objects are added to the archive.

The use of blockchain as main repository for digital archives could overcome all the described challenges. A blockchain is intrinsically distributed, thus all the information it stores is replicated on all the nodes of the network. This guarantees that stored objects are preserved over the time. In fact, if a failure occurs on a node of the network, information is not lost.

A blockchain guarantees also decentralization, in the sense that data are not hosted by a single central authority.

Finally, a blockchain overcomes the described security issues because of its intrinsic nature, based on cryptography.

It is worth mentioning, though, that identity authentication is performed by checking if a transaction is signed with a correct private key. In other words, identity is associated with key ownership, with no guarantees over the real identity of the owner of that key.

5. A POSSIBLE SCENARIO

Blockchain technology could be used to implement a platform for the registration of minor artworks. We refer to works that are artistically relevant 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. The aim would be to preserve artwork that, since it is often not well protected, can easily become subject to counterfeiting, be stolen or damaged by natural disasters. Examples of such works could be those kept in some small libraries or churches, or even in private households.

In details, the platform should guarantee the following aspects:

Protection: 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);

Anti-counterfeiting: protect artworks against forgery and allow correct identification of works in case of loss and subsequent discovery;

Integrity: make sure that the digital description of the artwork is not subject to unauthorized changes;

Traceability: trace all movements of individual artworks.

The previous objectives can be achieved through the implementation of a service based on blockchain. 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.

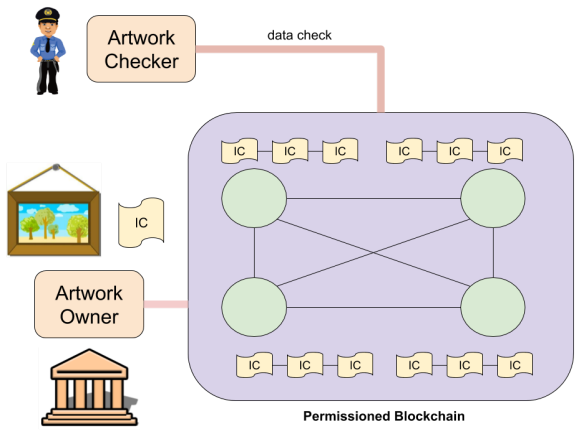

Figure 1: A possible implementation of the service for the registration of minor artworks.

In this implementation, each minor artwork is associated to an Identity Card (IC), which contains all the metadata about the artwork: of course its author, the date and place of creation, as well as its current owner and location, and all the other relevant information that is dependent on the specific kind of art, along with administrative and technical metadata.

An IC could be implemented through a smart contract having three main methods: `insertArtwork(data)` and `updateArtwork(data)`, which modify the blockchain, and the function `getInformation(artwork)` that accesses data stored in the contract.

The collection of all artworks is stored into a permissioned blockchain, which is composed of different nodes, hosted by some authorized organizations, such as governments and representative institutions. The fact that different institutions have a copy of the blockchain helps reducing errors and incompatibilities in the registration and management of the artworks.

Two types of users can access the blockchain of artworks: *artwork owners* and *artwork checkers*. Artwork owners host artworks (e.g. museums) and are responsible of inserting the IC of new artworks in the blockchain. In addition, artwork owners update ICs, whenever a change occurs, such as a temporary movement of the artwork for an exhibit or a restoration. Artwork checkers are responsible of verifying that artworks are not counterfeit or stolen. Examples of artwork checkers could be police departments or art experts.

6. REFERENCES

[1] ARMA International. *Generally Accepted Recordkeeping Principles*. 2017.

[2] Borghoff, U. M., R ¨odig, P., Scheffczyk, J., and Schmitz,L. *Long-Term Preservation of Digital Documents: Principles and Practices*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[3] Cachin, C. *Architecture of the hyperledger blockchainfabric*. In Workshop on Distributed Cryptocurrencies and Consensus Ledgers. 2016.

[4] ISO. *ISO 15489-1/2: 2016- Information and documentation - Records management*. 2016.

[5] Nakamoto, S. *Bitcoin: A peer-to-peer electronic cashsystem*.

[6] Schwartz, D., Youngs, N., Britto, A., et al. *The rippleprotocol consensus algorithm*. Ripple Labs Inc WhitePaper. 2014.

[7] Swan, M.Blockchain: *Blueprint for a new economy*. O'Reilly Media, Inc.. 2015.

[8] Wood, G. *Ethereum: A secure decentralised generalisedtransaction ledger.Ethereum Project Yellow Paper.* 2014.

[9] Zheng, Z., Xie, S., Dai, H.-N., and Wang, H. *Blockchain challenges and opportunities: A survey*. Work Pap. 2016.