InterPARES Trust Project Report

Title and code:		Model for Preservation of Trustworthiness of the Digitally Signed, Timestamped and/or Sealed Digital Records (TRUSTER Preservation Model) (EU31)
Document type:		Final report
Status:		Final version
Version:		1.3
Research domain:		Control
Date submitted:		3 February 2018
Last reviewed:		8 March 2018
Author:		InterPARES Trust Project
Writer(s):		Hrvoje Stančić
Research team:		Göran Almgren (Enigio Time AB), Hans Almgren (Enigio Time AB), Natasha Khramtsovsky (Electronic Office Systems LLC), Victoria Lemieux (UBC), Željko Mikić (TechEd), Elis Missoni (FINA), Hrvoje Stančić (FHSS), Mats Stengård (Enigio Time AB) FHSS GRAs: Andro Babić, Nikola Bonić, Vladimir Bralić, Hrvoje Brzica, Magdalena Kuleš, Anabela Lendić, Ksenija Lončarić, Ivan Slade Šilović, Ana Stanković, Ira Volarević

Document Control

Version history			
Version	Date	By	Version notes
0.1		H. Stančić	Initial draft
0.3	16.12.2017	Team members	Working draft
0.4	17.12.2017	Team members	Working draft
0.5	19.12.2017	Team members	Working draft
0.6	27.12.2017	H. Stančić	Working draft
1.0	07.01.2018	H. Stančić	Draft for consultation at EU31 level
1.1	12.01.2018	V. Bralić, N. Khramtsovsky, M. Stengård	Improvement suggestions
1.2	03.02.2018	H. Stančić	Final version
1.3	08.03.2018	H. Stančić	Minor corrections

Contents

1. Abstract **4**
2. Research team **5**
3. Background **6**
3.1. Digital records in the context of long-term preservation **6**
3.2. Digitally signed records **6**
3.2.1. Digital signatures 7
3.2.2. Digital signatures in the context of cryptography 7
3.2.3. Confidence in the identity of a person who digitally signs a document 8
3.2.4. Digital Timestamps 9
3.3. Blockchain **9**
3.3.1. Fundamentals of blockchain and DLT 10
3.3.2. Applications of blockchain technology 15
3.4. Relevant standards and legal frameworks **18**
3.4.1. ISO 15489 – Information and documentation – Records management 19
3.4.2. ISO 14721 – Open Archival Information System Reference Model 19
3.4.3. DSS – Digital Signature Standard 19
3.4.4. eIDAS Regulation 20
3.4.5. ISO/TC 307 – Blockchain and Distributed Ledger Technologies 20
3.5. Current approaches to long-term archiving and preservation of digitally signed records **21**
3.5.1. OAIS and TDR 22
3.5.2. CRL and OCSP 22
3.5.3. Archival Timestamp 23
4. Research questions **24**
5. Aims and Goals **25**
6. Methodology **26**
7. Findings **27**
7.1. Case studies **27**
7.1.1. Case study 1 – digitally signed retirement fund records 27
7.1.2. Case study 2 – digitally signed e-tax records 27
7.1.3. Case study 3 – digitally signed medical records, procurement and supplier contracts, official political decisions and minutes of meetings 28
7.2. TRUSTER VIP (Validity Information Preservation) Solution: TrustChain **28**
7.2.1. Introduction 28
7.2.2. The TrustChain model 29
7.3. Discussion **35**
8. Conclusions **35**
9. Products **37**
10. List of figures and tables **37**
11. References **38**
12. Appendix 1 – TRUSTER Preservation Model (EU31) – Case Study Questionnaire **41**

1. Abstract

Long-term preservation of digital records that are digitally signed or have a digital seal attached to them is a challenge for the archival profession. Such digital records are not easy to preserve, not only because of the constant technology advancements, but also because the certificates they rely on have limited duration. Apart from the three well known approaches to preservation of digitally signed records, namely 1) to preserve the digital signatures, 2) to eliminate the signatures, and 3) to record the trace of the signatures as metadata, this research proposed the model of a fourth approach – to register the validity of the digital signature in a blockchain.[1]

The team raised 13 research questions corresponding to the three aims and three goals of the study. The aims of the research study were: 1) to addresses an important archival issue of preservation of digital records in the cloud by utilizing new technological concepts, i.e. blockchain, 2) to build a model suggesting how to preserve the trustworthiness of the digital records with digital signatures, certificates, timestamps or seals added to them, and 3) to investigate the possibilities of revalidation of the expired digital signatures, periodical re-signing of digital records or renewal of timestamps, addition of archival timestamps, injection of additional (timestamped) proof of existence etc. The goals of the research study were: 1) to achieve results that could be used to draft and/or to improve regulatory frameworks, 2) to achieve results that could be used to draft and/or to improve internal organizational policies and procedures, and 3) to achieve results that are relevant for organization and development of trusted archival services relying on ingest of trustworthy records.

To achieve the stated aims and goals and to answer the research questions, the research was divided into five, sometimes overlapping, phases over 19 months (March 2016 – September 2017): 1) review of the relevant literature, 2) development of three case studies, 3) testing of various recordkeeping and archival preservation use cases, 4) development of the model for preservation of validity of digitally signed and timestamped records, and 5) writing of the final report.

The case studies (CS) showed that in the three surveyed institutions, holding digitally signed retirement fund records (CS1), digitally signed e-tax records (CS2), and digitally signed medical records, procurement and supplier contracts, official political decisions and minutes of meetings (CS3), there were no digital preservation actions taken yet, that there are records with the expired and soon expiring digital signatures, that the records are vital for certain important processes involving citizens, and that there is a need for development of digital preservation strategies and policies for preservation of the validity of digital signatures.

The research team developed the model called “TRUSTER VIP (Validity Information Preservation) Solution: TrustChain”. The TrustChain is modelled as a blockchain-based solution that can enable archival institutions (or others with such needs) to avoid having to periodically re-sign (or timestamp) all their archived, digitally signed records. The TrustChain achieves that by checking the record’s signature validity at ingest and, if valid, writing the signature’s hash (and some metadata) in the blockchain. Signature validity is checked by all or, if their number is sufficiently high, some (qualified majority) of the institutions participating in the distributed network of interconnected trusted institutions. If the signature is deemed valid the information is permanently stored in the TrustChain blockchain, i.e. in the distributed ledger.[2]

We envision TrustChain, a blockchain-based VIP solution, being maintained by an international alliance of archival institutions.

In addition to the case study questionnaire found in the Appendix 1, the research produced three case study reports and the blockchain bibliography as separate products as well as the list of blockchain-related terms and definitions that were included in the InterPARES Trust terminology database.

[1] For specific use cases it was also recommended to consider a possibility of shifting preservation obligations from source digitally-signed records to a trusted database formed using these records; while disposing of (already unusable) source records. Such a shift requires changes in legislation and/or regulations and might need approval by the National Archives.

[2] Alternatively, services of a trusted third party may be used instead of (or as a provider of) a blockchain solution. In any case, a solution should be underpinned by corresponding legislation and/or regulations.

2. Research team

Lead Researcher:	dr. Hrvoje Stančić, associate professor, Faculty of Humanities and Social Sciences (FHSS), University of Zagreb, Croatia
Project Researchers:	Göran Almgren, Enigio Time AB, Stockholm, Sweden
	Hans Almgren, Enigio Time AB, Stockholm, Sweden
	dr. Natasha Khramtsovsky, Electronic Office Systems LLC, Moscow, Russia
	dr. Victoria Lemieux, associate professor, University of British Columbia (UBC), Vancouver, Canada
	Željko Mikić, TechEd Consulting Ltd. Zagreb, Croatia
	Elis Missoni, Financial Agency – FINA, Zagreb, Croatia
	Mats Stengård, Enigio Time AB, Stockholm, Sweden
Graduate Research Assistants: (FHSS)	Andro Babić, March 2016 – present
	Nikola Bonić, March 2016 – October 2016
	Vladimir Bralić, March 2016 – present
	Hrvoje Brzica, March 2016 – present
	Magdalena Kuleš, June 2016 – present
	Anabela Lendić, March 2016 – present
	Ksenija Lončarić, March 2016 – April 2016
	Ivan Slade Šilović, August 2016 – present
	Ana Stanković, March 2016 – April 2016
	Ira Volarević, March 2016 – October 2016

3. Background

3.1. Digital records in the context of long-term preservation

Digital records can nowadays be created in two ways – they can be digitized from existing paper records or born digitally. Digitization in the broadest sense represents the transformation of an analogue signal into a corresponding digital form, and in a more narrow sense it represents the transformation of different materials into a digital form, turning them into a binary code in the form of a computer file.[3] Digitization splits the notion of preservation into two parts – the preservation of the information content or the information recorded in a document and preservation of the physical object, i.e. the medium that carries the information. The information content is digitized and saved separately from the physical object. (Stančić, Digitization of documents, 2000) It is important to note that every digitally preserved record should have its characteristics of authenticity, reliability, integrity and usability intact. (ISO 15489-1:2016 Information and documentation – Records management – Part 1: Concepts and principles, 2016). A trustworthiness of a record refers to its accuracy, reliability and authenticity. (InterPARES Trust Terminology Database) Archiving and preservation represent a unique challenge due to the long-term nature of the activities. The problem of long-term preservation and maintenance of digital information can be interpreted as preserving records so that the technology they are based on does not become obsolete. Digital objects require constant and continuous maintenance and depend on complex ecosystem of hardware, software, standards and legal regulations which are constantly changing, being amended or replaced. When compared to analogue records, digital ones face greater risk of decaying, the reason for that being primarily the fast pace of information technology development. Preservation of digital records is much more than preservation of a computer file – the goal is to enable access to the content while at the same time ensuring that their important characteristics are preserved.

3.2. Digitally signed records

The result of e-business and digital communication is the creation of an ever increasing number of digital documents and digital records which might also contain digital signatures or digital seals attached to them. Therefore, it is necessary to analyze the challenges of long-term preservation of such digital records.

While technically the same, the difference between digital signatures[4] and digital seals is that digital signature can be associated only with a natural person and the signing key must be under the sole control of the signatory with the aim to sign, while digital seal can be associated only with a legal person, the signing key must be under the sole control of the process assigning a seal with the aim to ensure integrity and origin. (What is an electronic seal?) (eIDAS, 2014)

In order to be preserved for the long-term digitally signed records also must have the basic characteristics of authenticity, reliability, integrity and usability which require a more complex approach to preservation compared to digital records which are not digitally signed or stamped. Just as there is a difference between the short-term and long-term preservation of digital records, there is a difference between the preservation of digital records which are digitally signed or sealed and those which are not. Digitally signed or sealed records contain one more level of complexity in the form of a digital signature or seal which makes their preservation more complicated.

Even though digitally signed records can be preserved for a longer period, they may lose their legal validity if the digital record cannot be validated or if it loses its property of non-repudiation. If an error occurs in the process of digital signature validation, the trustworthiness of the digital record becomes deprecated. This issue arises because re-validation of historical digital signatures usually is not

[3] Croatian Encyclopedia (Miroslav Krleža Institute of Lexicography, 2017)

[4] The terms *electronic signature* and *digital signature* are often used interchangeably to mean the same thing. However, in this report the term *electronic signature* will be used when referring to the signatures in which the identity of the signatory cannot be verified while the term *digital signature* will be used when referring to the signatures where the Certificate Authority (CA) confirms the identity of the signatory (except in the citations where the original terminology will be cited).

supported by common software and requires reliable long-term preservation of digital certificates, CRLs etc. and, in longer term, other elements of historical PKI infrastructure. If any of the elements of this system malfunctions, digital signature validation will fail. This is especially important when preserving records that contain advanced digital signatures. (Herceg, Brzica, & Stančić, 2015)

3.2.1. Digital signatures

A digital signature is a code created according to cryptographic principles using the Public Key Infrastructure connected to a digital object which serves as a proof that the object has not been tampered with, and in some cases can be used to authenticate the sender's identity. (Mihaljević, Mihaljević, & Stančić, 2015) The Electronic Signature Act, which was in force in Croatia until 7 October 2017, defines digital signature as data in an electronic form added or logically connected to other data in an electronic form which serve to identify the signatory and the trustworthiness of the signed electronic document.[5] The EU eIDAS Regulation which superseded the Electronic Signature Act defines an electronic signature as data in electronic form which is attached to or logically associated with other data in electronic form and which is used by the signatory to sign.[6] Therefore, a digital signature represents the basic technology used to check a digital document's authenticity and a method of document content or electronic communication protection, whereas the digital certificate, which it relies on, makes it possible to confirm the identity of a natural or legal person. Thus, the basic requirement for a digital signature is the validation or confirmation of originality or authenticity of the signatory and the content of the signed communication. (Katulić, 2011)

Although a digital signature functions on the same principles as a traditional, wet ink signature there is an important difference between the two. While signing many different paper documents it is generally expected that the wet ink signature is the same on all documents.[7] On the other hand, while signing many different digital documents the digital signature will be a different binary string but it will still be associated with the document and the signatory. In other words, a person signs each digital document with a different digital signature because a digital signature is sent attached to a message in the form of a binary string. If the same string were used for multiple documents, anyone who received a digitally signed document could simply copy the string and attach it to another document, thus forging someone else's signature. Apart from the basic electronic signatures which the European Telecommunications Standards Institute (ETSI) defines as essentially the equivalent of a hand-written signature, with data in electronic form being attached to other electronic subject data as a means of authentication, there are also *advanced electronic signatures*. According to the eIDAS Regulation, whose definition is based on the Directive 1999/93/EC, which has since been superseded, the *advanced electronic signature* must meet the following requirements: "a) it is uniquely linked to the signatory; b) it is capable of identifying the signatory; c) it is created using electronic signature creation data that the signatory can, with a high level of confidence, use under his sole control; and d) it is linked to the data signed therewith in such a way that any subsequent change in the data is detectable."[8]

3.2.2. Digital signatures in the context of cryptography

To understand the principles of cryptographic security in digital signatures and digital certificates it is important to observe the environment in which the digital signature came about and the pre-existing infrastructure and methods of document security it used.

The word cryptography stems from the Greek word κρύπτω ('kripto') meaning hidden or secret and γράφειν ('grafein') meaning to write. Cryptography deals with the study of message sending methods, in which the message may only be read by the intended recipient of the message. The goal of cryptography is, therefore, to enable two persons to communicate through an insecure channel of

[5] Electronic Signature Act – NN 010/2002 (Croatian Parliment, 2002)
[6] eIDAS Regulation (European Parliament, 2014)
[7] In some legislations there is no such expectation (e.g. in the USA). And in Russia there is no official requirement of uniqueness of wet signature – it's more common practice than law. There are contexts in which different wet signatures are routinely used by the same person – e.g. bank clerks are using simplified (sometimes entirely different) wet signatures for signing transaction documents.
[8] eIDAS Regulation (European Parliament, 2014)

communication, so that a third person cannot understand them. The person sending the message is called the *sender*, the person receiving the message the *recipient* and the third person trying to intercept the message is called the *attacker*. Using a key agreed upon beforehand, the sender transforms the message. The original message is referred to as *plaintext*, the transformation procedure *encoding* or *encryption*. The result of the encryption is called a *ciphertext*. The sender sends the ciphertext through an unsafe channel of communication. Even if the attacker intercepts the message, (s)he cannot decipher or understand it without knowing the key. The key is in fact a mathematical function which is used for encryption and decryption. A cryptosystem is thus formed by the message, the ciphertext and the key which holds the encryption and decryption information. (Ibrahimpašić & Liđan, 2011)

Cryptosystems can be a) symmetrical – those which use the same key for encryption and decryption (meaning that the sender and recipient must safely exchange keys), and b) asymmetrical – those which use a combination of a public and secret (private) key (what one key of the key pair encrypts, the other key of the same pair, and only it, can decrypt and vice versa). The latter, asymmetrical cryptosystems, form the basis for the Public Key Infrastructure (PKI) and digital certificate systems.

Digital signatures use a combination of the public and private keys so that any person looking at a digitally signed document may ascertain whether the person indicated in the digital signature truly signed a document. This process is referred to as *authentication*. Because the keys serve as proof of authenticity, in theory the signatory cannot deny that (s)he has indeed signed the document. This principle is known as *irrevocability* or *non-repudiation*. (Ibrahimpašić & Liđan, 2011)

3.2.3. Confidence in the identity of a person who digitally signs a document

The problem with the public key cryptosystem is the problem of a secure connection between the key and the person, i.e. the question of the identity of the person who digitally signs a document remains open. Successful signature verification does not mean that the document was signed by the indicated person, but only that it was signed with the secret key that corresponds to the public one. Therefore, in the public key cryptosystem, there is only confidence in the successful key exchange but the true identity of a person cannot be trusted. The solution to that is the PKI system, i.e. the public key infrastructure, where the trusted third party (certification service) guarantees the identity of a person and the connection of that person to the pairs of private and public keys. This technology is based on the ITU-T X.509 recommendations from 1988[9] and the RFC 3280 from 2002[10]. The value of such a system is in its flexibility to provide services and applications for identification, authentication, digital signatures as well as security and secrecy. PKI is actually a system of digital certificates, certification and registration services that check the identity of the user, and this is its primary purpose.

The PKI system can be used to enable:

- identity verification,
- integrity of information,
- safer data exchange processes,
- public access to state and other e-services,
- acceptance of various electronically filled documents, and
- secure communication with employees at remote locations.

The PKI system provides the components necessary for managing (issuing, verifying and revoking) public keys and certificates (as well as their storage and preservation). It also provides secure authentication of communication participants, exchange of documents with the possibility of encryption, digital signing and co-signing and a unique registry of public keys in the form of a digital certificate.

[9] Information technology – Open Systems Interconnection – The Directory: Public-key and attribute certificate frameworks. Current version is Edition 8, 10/2016, http://www.itu.int/itu-t/recommendations/rec.aspx?rec=X.509

[10] Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile. Current version is RFC 5280 which was updated by RFC 6818, https://tools.ietf.org/html/rfc5280

The basic element of this system is a *digital certificate*. It is used for more demanding public key encryption implementations. A digital certificate confirms the connection between a secret key owned by a particular person and the associated public key. It is a system where the identity of the person is stored together with the corresponding public key, and the entire structure is digitally signed by a trusted third party (certification service). A digital certificate is issued for a limited period. The validity of the certificate can be revoked, i.e. the certificate may be revoked, even before the end of the period for which it was originally issued. The validity of a digital certificate can be verified by checking the digital signature, but there must be a direct trust or a trust chain to the certification authority certifying the digital certificate. The format of the digital certificate is defined by the third version of the X.509 standard.

The next element is the Certificate Authority (CA). It issues and revokes digital certificates, manages them, keeps them and guarantees their validity. In this system, CA is a trust entity and a third party. Registration Authority (RA) can also participate in the process of issuing certificates. It handles users' requests to issue digital certificates, registers users, and cooperates with CA during the issuance of certificates. RA ensures the correct physical identification of users, thus ensuring the non-repudiation characteristic of digital signatures. Regardless of whether RA is used or not, CA is responsible for issuing digital certificates. In addition to RA and CA, there is a Certificate Repository (CR) where public keys, user certificates and Certification Revocation Lists (CRLs) are stored. As already mentioned digital certificates have a certain period of validity (usually two to five years) and may be terminated or revoked if the user or certificate has been compromised. There are two ways of knowing whether a digital certificate is revoked. The first way is to check whether the certificate revocation information has been published on the CRL. The second way is by using the OCSP (Online Certificate Status Protocol) – an internet protocol which is used for obtaining the revocation status of a certificate[11].

3.2.4. Digital Timestamps

In the context of digital signatures, the *digital timestamp* plays an important role. It represents a digitally signed certificate of a timestamp issuer which confirms the existence of the data, documents or records to which the timestamp relates, at the time stated on the timestamp. The digital timestamp ensures reliable proof that the data, document or record originated earlier or just before the time indicated in the digital timestamp. Any subsequent changes to data, documents, records or timestamp are not allowed and can be easily detected. Therefore, the digital timestamp confirms: 1) that the data, document or record at hand existed in that form at the time indicated in the timestamp, 2) that the data, document or record was not changed after the time indicated in the timestamp The Timestamping Authority (TSA) digitally signs the hash value of the data, document or record along with the time value (coming from a trusted source, e.g. it can be linked to Coordinated Universal Time) thus issuing digital timestamp which is subsequently combined with the data, document or record and the signatory's private key to create the digital signature with the indication of the time of signing.

3.3. Blockchain

One of the possible solutions for the long-term preservation of digitally signed records is the blockchain technology. Blockchain is best known as the technology in the background of digital currencies (cryptocurrencies) which is already being applied in various other areas for a variety of purposes. This technology is by its very nature an archiving technology because everything that is being recorded can no longer be changed or deleted.
Blockchain represents a distributed database of (transaction) records storing hash values of data, information, transactions, documents or records. The blockchain is associated with the concept of Distributed Ledger Technology (DLT). The term itself consists of two terms – the term "block" refers to the complete set of contents, and the term "chain" refers to the interconnection of the blocks to each other. This chain grows linearly, and the creation of a new block, in the broad context of cryptocurrencies, is called *mining*.

[11] Defined by RFC 2560 – X.509 Internet Public Key Infrastructure Online Certificate Status Protocol – OCSP from 1999. Current version is RFC 6960 from 2013, https://tools.ietf.org/html/rfc6960

3.3.1. Fundamentals of blockchain and DLT

In order to better understand the blockchain and distributed ledger technologies one need to understand the underlying technologies and concepts. Therefore, hash algorithms, Merkle tree, distributed consensus, and finally, blockchain will be explained next.

3.3.1.1. Hash algorithms

Hash, or message digest, is a one-way function that calculates a unique fix-length string out of any data, information, document or record of any size. The *one-way* characteristic means that it is not possible to recreate the original document by knowing its hash. It should be practically impossible during the set period of time to create "collisions", i.e. to have two or more meaningful records with the same hash value (produced by a given hash function). The resulting hash value is also referred to as *digital fingerprint*. Therefore, it is said that the hash algorithm is *collision resistant*. Another characteristic of the hash algorithm is that it is *pseudorandom*. It means that it is unpredictable but also that even if, intentionally or unintentionally, only one zero is changed to one in the binary stream, the resulting hash will be significantly different from the original one. This also means that the hash function is *deterministic*, i.e. that it will always result with the same hash out of the same input data. (Drescher, 2017) It is more secure than the CRC (Cyclic Redundancy Check) because it is possible to manipulate the data without the change in previously generated CRC or, vice versa – it is possible to change any file's CRC to any value.[12]

There are many different hash functions, e.g. Adler32, Haval, LM, MD, NTLM, RipeMD, SHA (Secure Hash Algorithm), Snefru, Tiger, Whirlpool etc. Some of them can be of different strengths, i.e. different string lengths, e.g. MD2, MD4, MD5 or SHA-1, SHA-224, SHA-256, SHA-384, SHA-512 etc.[13] Figure 1 shows the different hash values of the plaintext of the title of this research.

Original text	Model for Preservation of Trustworthiness of the Digitally Signed, Timestamped and/or Sealed Digital Records (TRUSTER Preservation Model)
Original bytes	4d6f64656c20666f722050726573657276174696f6e206f66... (length=137)
Adler32	ae4831e2
CRC32	ed29d1fa
Haval	15a4971d923ad4a7adce708c83fbb512
MD2	f16faa8a18b6695714cfc34ea3277c51
MD4	d51acfc5b54f21f225fe82819ddcfc21
MD5	9b3a1431647b127f5d133f0c54adb0f5
RipeMD128	ad9f290a21c91df7a9d5460bf800fc0a
RipeMD160	069b84458115557d3025b308d3663925979cf36d
SHA-1	2f8a2914921342addb47ed18b89e0b2104a113c0
SHA-256	064bfb90c5e6587bbd8a55122f8a53daaaa79da1b5b9ef6108072f4bdf83ab65
SHA-384	3aca9c561d4be7db34c651377699ac40e1521be9c8b0eebb92e2e2a51f0bcae589563a1b34f4a00dba3c68bd0005fb94
SHA-512	8380e7138bbadf22372b9326742bf9c3fee10f1a1b2d9b133abfc45292a976bdb1bcb7fbede132c87aaab90b5bdd4a08cf5c97d960c16499d8b7eef99771ee15
Tiger	ca2d84335f6ee13101f69dcccad0cbe79a6621d45580f5db
Whirlpool	c8cad58472c1c913b1b15ae1d58bb61749776daa4c75842ac2bff15e7d8312e83b34535e85f01e1275eed4751bdd96f4351c5e87344852792aeefabd6300ef7e

Figure 1. Example of hash values

Figure 2 shows two significantly different hash values resulting from the application of the same SHA-256 hash algorithm for two plaintext inputs that differ only in one sign – semicolon is used instead of comma.

[12] Forcing a file's CRC to any value (Nayuki, 2016).

[13] Some of the mentioned hash functions have been outdated for a long time and are included here for historical and demonstration reasons.

Original text	Model for Preservation of Trustworthiness of the Digitally Signed, Timestamped and/or Sealed Digital Records (TRUSTER Preservation Model)
SHA-256	064bfb90c5e6587bbd8a55122f8a53daaaa79da1b5b9ef6108072f4bdf83ab65
Original text	Model for Preservation of Trustworthiness of the Digitally Signed; Timestamped and/or Sealed Digital Records (TRUSTER Preservation Model)
SHA-256	ec6ab344f1f433b6f680ef71a3bef0053deff852e57abb7725158889e7b2d791

Figure 2. Example of pseudorandom characteristic of hash function

3.3.1.2. Merkle tree

Hash values may be hashed together. This will be illustrated by the following example (see Figure 3). A small company creates 10 documents in the morning and 10 documents in the afternoon. A hash value is calculated for each document. At noon, all 10 hash values from the morning are hashed together to get just one "morning" hash. At the end of the day, for example Monday, all 10 afternoon hash values are again hashed together to get one "afternoon" hash. After that, "morning" and "afternoon" hash values are hashed together to get one hash value for Monday. This hash is called *root hash* or *top hash*. This example is further extended in Figures 5 through 9 explaining the blockchain concept.

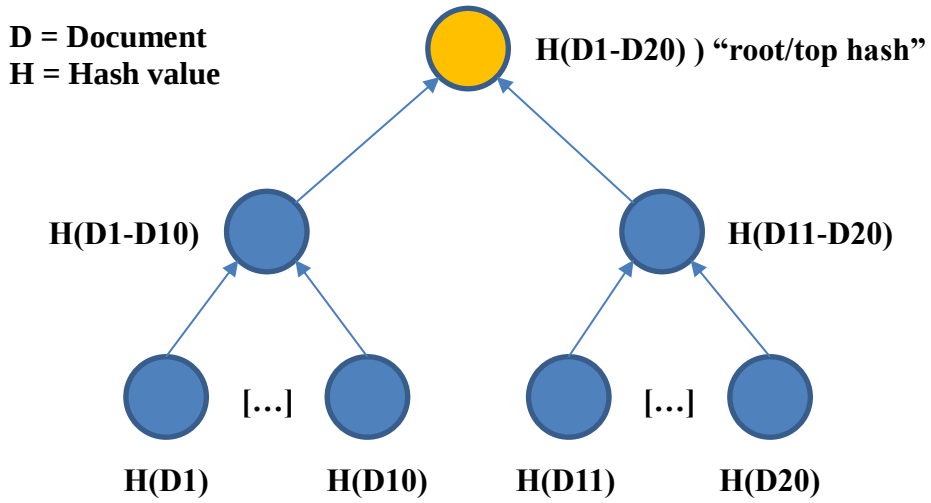


Figure 3. Merkle tree

This approach was first introduced in 1980 by Ralph C. Merkle. (Merkle, 1980) Since the structure resembles the tree structure (upside-down) it was named Merkle tree. Before explaining the concept of blockchain which utilizes the Merkle tree approach, the concept of distributed consensus will be explained.

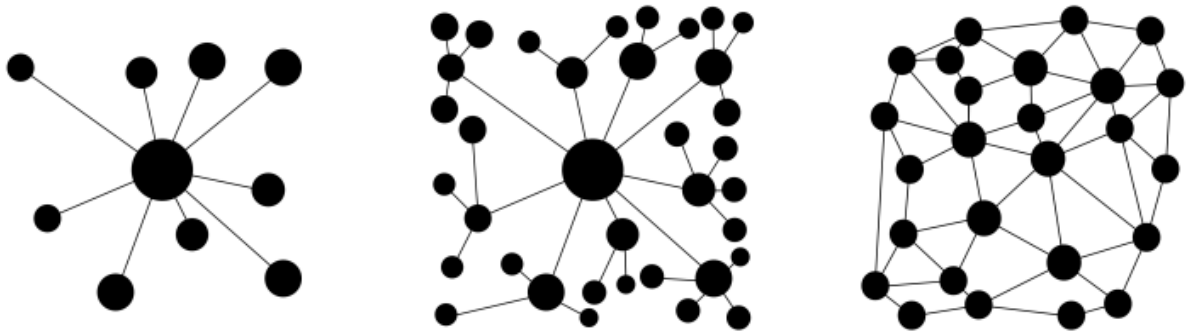
3.3.1.3. Distributed consensus

Blockchain is using distributed (peer-to-peer) network. Basically, there are three types of network topologies – centralized, decentralized and distributed topology[14] (Figure 4). The centralized network has one central server to which other computers are connected. This type of network topology has a centralized control and a single point of failure (central server). Hackers could target the central server. The decentralized network has several servers to which other computers are connected. This type of network topology has decentralized control and is more secure than the centralized network but servers can still be identified and attacked. The distributed network has no center(s) since all interconnected computers are treated equally. This type of network topology has no single point of control and therefore no single point of attack.

[14] ISO groups recognized the difference between physical architecture and control/governance architecture. Centralized system may have redundancy and distributed elements, e.g. there may be several copies of the main control server situated in different geographical locations.

The first two types of network topology could also be seen as applied to the structure of government, institutions or organizations. There are central offices, branches etc. This also means that the power is more or less centralized, or that the trusted third party in communication or exchange is needed. For example, if a person A wants to send certain amount of money to a person B, at least person B has to have a bank account. The bank in this example is the trusted third party. On the other hand, by application of the distributed network concept one could avoid the need for the trusted third party and exchange the money directly.

This is possible by applying the principle of distributed consensus in which every participant (node) checks every event in the ledger ("main book"/database). Consensus is used to determine the truth, i.e. to prevent double spending of the same amount of money or two different instances of the same record. The event (e.g. monetary transaction or registration of a record) is valid only if the qualified majority (50%+1 node) agrees upon it.



Centralized network | Decentralized network | Distributed network

Figure 4. Three types of network topology[15]

3.3.1.4. Blockchain concept

The Merkle tree approach was used by Satoshi Nakamoto for creating virtual/crypto-currency Bitcoin (Nakamoto, 2008) which in turn sprung wider application of the blockchain technology.
The blockchain creates a chain of linked blocks. This will be illustrated by extending the example explaining the Merkle tree and shown in Figure 3. The previously mentioned company can repeat the Monday hashing process on Tuesday. This will result with two hash values – one for each day. Those two values could further be hashed together to create a new single top hash uniting single hashes from Monday and Tuesday. This single hash value would be further combined with the Wednesday hash value to create a new top hash etc. Each new top hash is calculated from the day's hash and a previous top hash thus linking the top hashes (Figure 5).

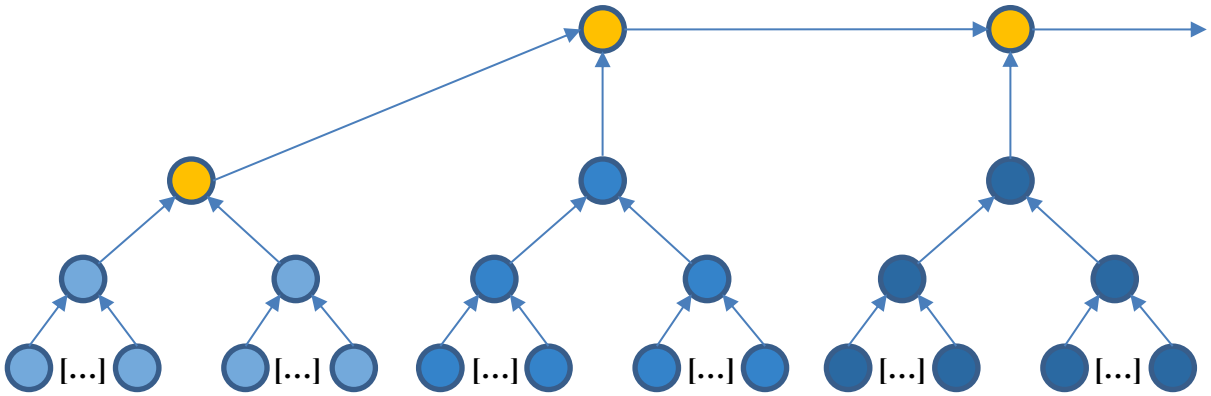


Figure 5. Linking of the hash values

[15] Image source: http://bluenetworks.weebly.com/syngeneia-in-the-history-of-pergamon.html

Blockchain builds on that and asks all the participating nodes to confirm the creation of the new top hash. As per distributed consensus principle, a new block is confirmed when the qualified majority agrees upon it (Figure 6). The cryptocurrencies also implement the time-consuming computational tasks called *hash puzzles* in order to calculate the new block's hash, i.e. to implement the concept of *proof of work* and consequentially create the value of the cryptocurrency which is originating from the computational resources and time used. However, the explanation of that process is out of scope of this report.

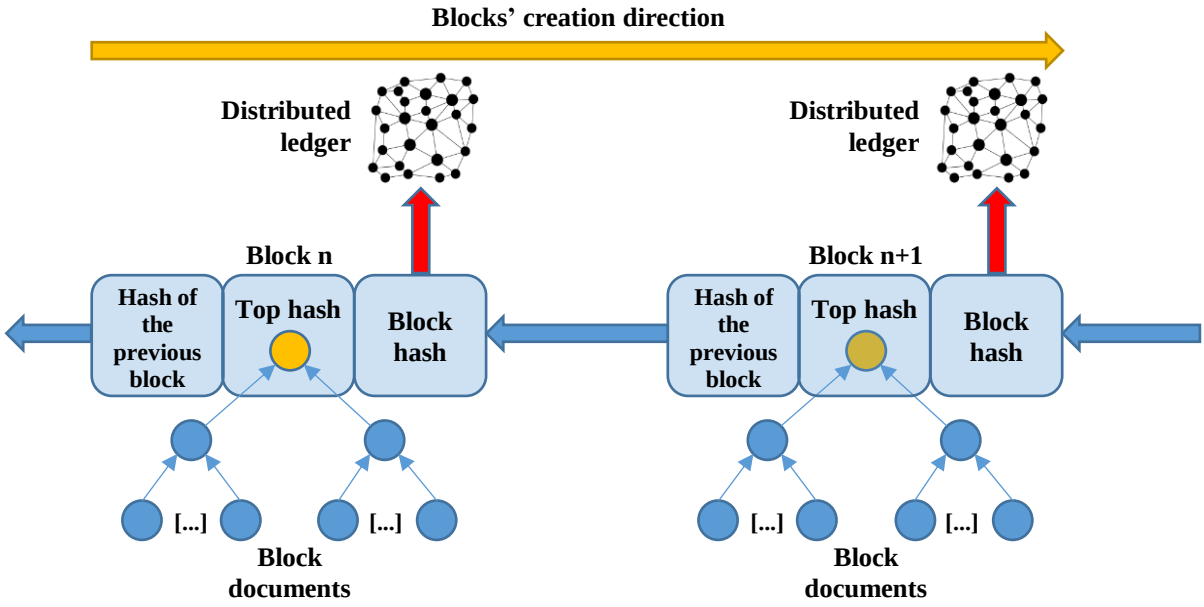


Figure 6. Blockchain creation

It is also important to stress out that one block can have many hashes (events, transactions) that are blocked together (Figure 7). New blocks, when confirmed, are timestamped and recorded by every participating node thus creating the distributed ledger. New blocks are created in regular time intervals. Depending on how time critical the creation of new blocks is, the time intervals can range from, e.g. every 10 or 15 minutes to every minute or, when the latency is important, it can be scaled down to every second.

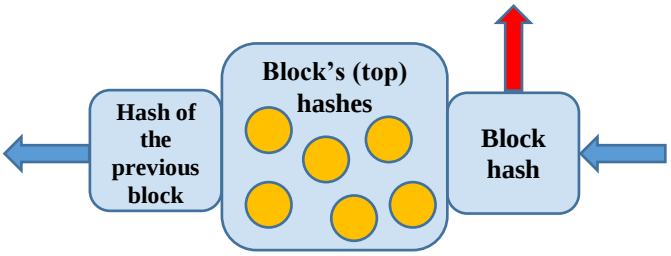


Figure 7. Multiple hashes combined in one block

There are several strong sides of the blockchain concept. Firstly, although it is possible to store data on the blockchain, in this example only hashes are stored (registered). The actual data, documents or records being hashed are stored in the institutional system and remain under institutional control. Secondly, each additional block reinforces the preceding ones since the chain is formed of the linked blocks. Thirdly, any attempt to modify a block would invalidate all subsequent blocks and would be easily detected (Figure 8). Therefore, no changes to the already created blocks are possible.

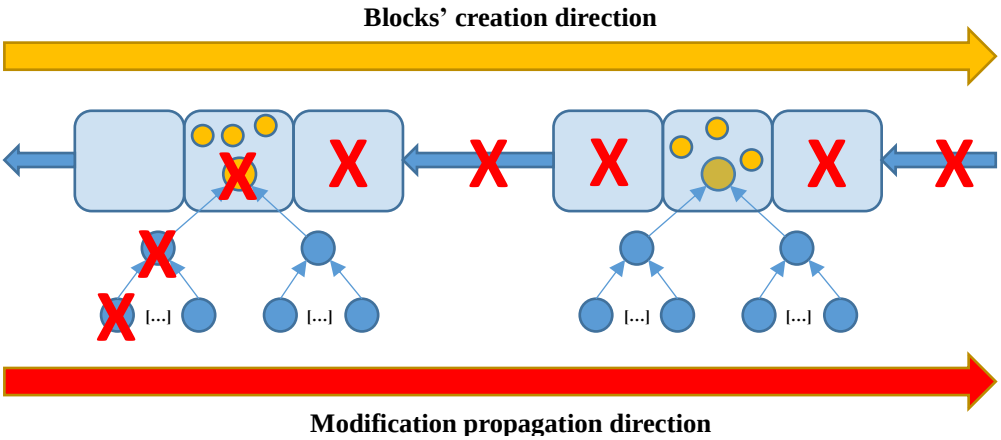


Figure 8. Hash modification propagation through the blockchain

The blockchain contains proof that a hash, and in that sense that a data, document, record or transaction, was part of the original set of hashes the chain was built upon (Figure 9).

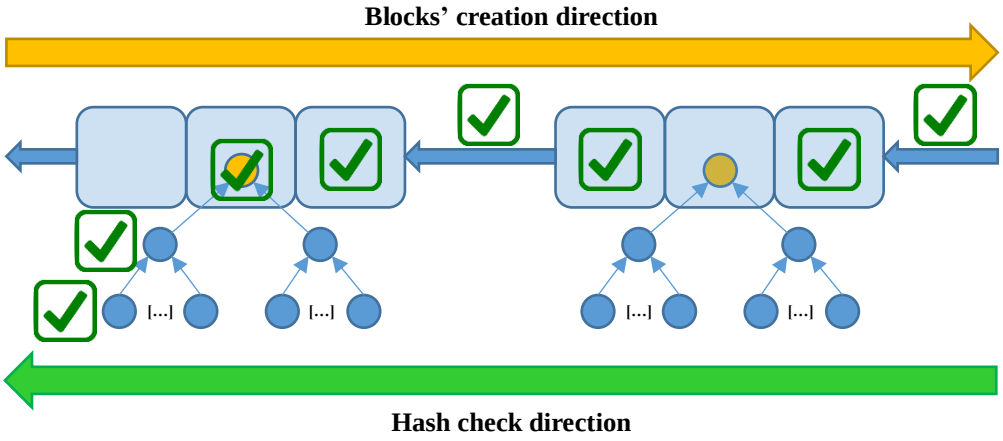


Figure 9. Checking of a hash value in the blockchain

It should be noted that by obtaining control over the qualified majority of the interconnected nodes, the attacker might obtain control of the new block creation. However, in order to change a hash value that has already been logged on the blockchain, the attacker should recalculate all subsequent block hashes. This might be a shortcoming if the blockchain is used without the proof of work while if it is used the computational and time costs outweigh the benefit. Nevertheless, the TrustChain model developed in this research study tackled the problem of using blockchain without the proof of work (see Chapter 7.2.).

The problem of weakening of the cryptographic algorithms used to create the original hash values with time might be solved by hashing the first next block with a stronger hashing algorithm and continue to use it. The other solution is to reach a consensus of qualified majority of nodes to cut off the chain at the point when the stronger algorithms are to be used thus starting the new chain. The old part of the

chain should be safely preserved by all nodes and might be wrapped with a stronger algorithm hash for security reasons.

For any distributed blockchain it is critical to reach a critical size, i.e. to involve a significant number of nodes in order to gain security against the potential attacks. What is a *significant number* depends on the purpose of the blockchain implementation and the attractiveness for the possible attackers. In this context it should be mentioned that there are two possible types of the blockchain implementation – a public and a private blockchain. Table 1 summarizes the differences between the two.

Table 1. The differences between public and private blockchain

Public blockchain	Private blockchain
Anyone can freely write data without permission granted by any authority	Only known and trusted (authorized by an authority) participants can freely write data
No point of control	No point of control (except initial authorization)
(Relative) anonymity	No anonymity
Examples: Bitcoin, Ethereum	Example: A group of partnering archives

What is lacking from the archival perspective up until now is the archival bond. Lemieux and Sporny realize that "even though the time-ordered nature of the transactional records is preserved, the link to their procedural context, and relationship to other transactional records relating to the same procedure, is not." They propose that "through use of ontologies to represent the procedural context of ledger entries, it is possible to instantiate the archival bond between ledger entries as records of a variety of transactions." (Lemieux & Sporny, 2017)

Taking in account all characteristics of the blockchain, and its underlying technologies and concepts, it could be concluded that the blockchain can be used to:

- confirm integrity of a record,
- confirm that a record was existing or created at a certain point in time (i.e. not after it was timestamped and registered in the blockchain),
- confirm sequence of records,
- support/enhance non-repudiation of a record,
- improve the validation possibilities of digitally signed records during the long-term preservation, and
- ensure that digital processes cannot be manipulated (see Chapter 7.2.).

3.3.2. Applications of blockchain technology

"Blockchain technology has attracted attention as the basis of cryptocurrencies such as Bitcoin, but its capabilities extend far beyond that, enabling existing technology applications to be vastly improved and new applications never previously practical to be deployed. Also known as distributed ledger technology, blockchain is expected to revolutionize industry and commerce and drive economic change on a global scale because it is immutable, transparent, and redefines trust, enabling secure, fast, trustworthy, and transparent solutions that can be public or private. It could empower people in developing countries with recognized identity, asset ownership, and financial inclusion." (Underwood, 2016) There are a lot of blockchain applications that could transform society – some of them being blockchain-based financial services, smart property applications (e.g. registration of title to assets), smart contracts, applications in the healthcare or music sectors, notarization, tracking of provenance as well as the e-Government applications like public voting, identity management etc. Further, application of blockchain in the public voting and in the healthcare sector will be explored.

3.3.2.1. Blockchain and DLT in public voting systems

Following the enormous success of Bitcoin and other blockchain-related technologies and their impact on many sectors outside financial, a question could be raised if the same blockchain technology can be applied to help the modern democratic processes, still relying mainly on paper (Bradbury, 2014). "Blockchain can serve as the medium for casting, tracking, and counting votes so that there is never a

question of voter-fraud, lost records, or fowl-play. By casting votes as transactions within the blockchain, voters can agree on the final count because they can count the votes themselves, and because of the blockchain audit trail, they can verify that no votes were changed or removed, and no illegitimate votes were added." (Huminski, 2017)

European Parliament (Boucher, 2016) emphasizes how, since the turn of the century, e-voting has been considered a promising and (eventually) inevitable development, which could speed up, simplify and reduce the cost of elections. Now we can continue trusting central authorities to manage elections or to use blockchain technology to distribute an open voting record among citizens. Many experts agree that e-voting would require revolutionary developments in security systems. The debate is whether blockchain will represent a transformative or merely incremental development, and what its implications could be for the future of democracy. Swan (Swan, 2015) argues that it is possible to apply the idea of using blockchain technology to provide services traditionally provided by the State in a decentralized, cheaper, more efficient, and personalized manner.

Although a comprehensive discussion of possible blockchain-based models of governance does not yet exist at academic level, some authors (Atzori, 2016) have proposed possible applications of the blockchain technology in the governmental sector. Based on Swan (Swan, 2015), Atzori (Atzori, 2016), among others, summarizes the main principles and problems of governance that blockchain could mitigate:

> **Centralized government.** Throughout history, centralized political organizations like state, bureaucracy and representative democracy have been a solution to a scaling problem. A centralized authority in any hierarchical organization can be defined in computer terms as a Single Point of Failure (SPOF): if its functioning is not optimal, the whole system and its participants will be negatively affected by it. Centralized vertical authority has become the main organizational model in society, simply because there has not been a better alternative so far. For the first time in history, citizens can now reach consensus and coordination at global level through cryptographically verified peer-to-peer procedures, without the intermediation of a third party. The blockchain technology can lead towards a new era of decentralization on a large-scale, in which human factor is minimized. In this scenario the decentralized government software could be used as a collaborative platform for DIY governance using which anyone could create his/her own government services (in a "blockchain nation"). Many new decentralized governance models and services can therefore be implemented and experienced through the blockchain (Swan, 2015).
>
> **Power of individuals**. While the state bases its action on coercion, the blockchain can provide governance services in a more efficient and decentralized way, without having to relying on force. This allows a more horizontal and distributed diffusion of authority, in which the source of legitimacy are the individuals themselves. Using the blockchain as a permanent, encryption-secured public record repository, human agents as representatives can be replaced by smart contracts and Decentralized Autonomous Corporations (Swan, 2015). The blockchain technology allows more granular and personalized government services. Using the blockchain as a permanent public records repository, it is possible to store all government legal documents, such as contracts, identification cards, passports, lands deeds, etc. in a cheaper, more efficient and decentralized way. Anyone can create a private blockchain and a decentralized do-it-yourself-governance system (Swan, 2015).
>
> In order to implement these two principles on a blockchain based technologies, one of the key areas is developing systems to make democracy more effective, but also more transparent at the same time. To achieve this goal, there is a need for decentralized voting systems.
>
> The blockchain protocol is a means of logging and verifying records that is transparent and distributed among users. Usually, votes are recorded, managed, counted and checked by a central authority. Blockchain-enabled e-voting (BEV) would empower voters to do these tasks themselves, by allowing them to hold a copy of the voting record. The historic record could then not be changed because other voters would see that the record differs from theirs. Illegitimate votes would be much harder to add, because other voters would be able to scrutinize whether votes were compatible with the rules (perhaps because they have already been counted, or are not associated with a valid voter record). BEV would shift power and trust away from the central actors, such as

electoral authorities, and foster the development of a tech-enabled community consensus (Boucher, 2016).

Swan (Swan, 2015) gives an overview of three possible approaches in creating such systems:

1) The Liquid Democracy system. In the Liquid Democracy system, a party member can assign a proxy vote to any other member, thereby assigning a personal delegate instead of voting for a representative. The idea of delegated decision making, supported and executed in blockchain-based frameworks might have wide applicability beyond the political voting and policy making context. Ideas for a more granular application of democracy have been proposed for years, but it is only now with the Internet and the advent of systems like blockchain technology that these kinds of complex and dynamic decision-making mechanisms become feasible to implement in real-world contexts.

2) Random-Sample Elections. In addition to delegative democracy, another idea that could be implemented with blockchain governance is random-sample elections. In random-sample elections, randomly selected voters receive a ballot in the mail and are directed to an election website that features candidate debates and activist statements. As articulated by cryptographer David Chaum (Chaum), the idea is that (like the ideal of a poll) randomly sampled voters would be more representative (or could at least include underrepresented voters) and give voters more time to deliberate on issues privately at home, seeking their own decision-making resources rather than being swayed by advertising. Blockchain technology could be a means of implementing random-sample elections in a large-scale, trustable, pseudonymous way.

3) Futarchy: Two-Step Democracy with Voting. Another concept is *futarchy*[16], a two-level process by which individuals first vote on generally specified outcomes (like "increase GDP"), and second, vote on specific proposals for achieving these outcomes. The first step would be carried out by regular voting processes, the second step via prediction markets. As with random-sampling elections, blockchain technology could more efficiently implement the futarchy concept in an extremely large-scale manner (decentralized, trusted, recorded, pseudonymous). There is the possibility that voting and preference-specification models (like futarchy's two-tiered voting structure using blockchain technology) could become a common, widespread norm and feature or mechanism for all complex multiparty human decision making.

Although there have been significant developments on how the blockchain could be implemented in the government sector, there is still much debate. The proposed systems such as BitCongress (Rockwell) combine old world democratic concepts, blockchain technology and Bitcoin as the underlying backbone. The main issue here is guaranteeing vote integrity from end to end – where the blockchain may be a useful means of guaranteeing vote integrity at the back end. The key question is how to ensure widespread trust in the security and legitimacy of the system. As with the paper-based elections, it is not enough for the result to be fair and valid. The whole electorate, even if they are disappointed with the result, must accept that the process was legitimate and reliable. As such, beyond providing actual security and accuracy, BEV must also inspire broad public confidence and trust. Because the blockchain protocol is quite complicated, this may be a barrier to mainstream public acceptability of BEV. (Boucher, 2016)

3.3.2.2. Blockchain and DLT in healthcare sector

"Healthcare institutions suffer from an inability to securely share data across platforms. Better data collaboration between providers means higher probability of accurate diagnoses, higher likelihood of effective treatments, and the increased ability of healthcare systems to deliver cost-effective care.

Blockchain can allow hospitals, payers, and other parties in the healthcare value-chain to share access to their networks without compromising data security and integrity. (...) Blockchain would allow the hospital to tie patients to their data rather than tie them to their identity." (Huminski, 2017)

Internet technologies and devices are becoming more relevant in all aspects of human lives, including healthcare. Devices can be applied to help diagnose illnesses and save human lives. Also, health

[16] The name "futarchy" comes from government by futures markets. (Hanson, 2000)

records are expected to exponentially grow in the coming years as more data is collected daily. The data growth will be led by the development of new methods of diagnosis and analyzing devices. Therefore it is necessary to analyze which technologies are suitable for use in the near future in order to ensure secure connections and data storage. To use all the benefits of technological advancement in healthcare, large databases are created and applications developed which can use data for research, comparison and analysis in healthcare prediction, therapy, diagnosis or disease prediction. The White House National plan for the future of Artificial Intelligence (AI) in the domain of healthcare, and their recommendations for specific actions by Federal Agencies and other actors, determine the healthcare regulations in the domain of public safety and security. Their recommendations are largely concerned with the administration of big data and privacy. (White House – National Science and Technology Council, 2016) In the context of this technology numerous potential problems arise, for example what if a malevolent individual changes someone's medical record stating that one is not allergic to penicillin when one in fact is. During the next hospital treatment the patient could receive penicillin and die due to the data breach. Another example would be controlling a pacemaker or even insulin injector that is connected to the referent doctor or hospital in order to send parameters. If this data could be intercepted and manipulated, patients could be induced with a stroke, their pacemaker shut down or they could be injected with a higher dose of insulin.

When it comes to technologies used in healthcare, there are different products and devices on the market. An interesting device is a brain implant that can help damaged parts of the brain to be restored or increase the memory capabilities. (Drummond, 2010) Devices that can be connected to smartphones in order to send medical data using the internet or mobile data have been on the market for several years now, and their number is growing daily. Also, the market for and the functions of smartwatches is growing as is their potential use in healthcare monitoring and data collection. Devices which measure pulse, blood pressure, skin changes, ECG, EEG, and even mobile DNA testing devices for smartphones are available on the market nowadays.

One of the important topics is the security in a communication between new medical devices and medical databases. As of yet, this communication is not secure enough to ensure patients' safety. In order to enable the existing technologies to be used in vital and basic healthcare tasks, further research in the field of highly secure communication is required. Because the topic essentially deals with information security, the best available methods for this type of indirect diagnosis come from the financial sector.

However, for this technology to be implemented successfully, securing connections and record safety is crucial. Combatting this issue, the US government sponsored a contest in 2016 whose goal was to find the best solution for using blockchain technology in healthcare. One solution, developed as part of the contest, was the idea of using mobile data to monitor a person's health status by sending data via a special application to their referent doctor, while another solution dealt with database and record data security. Estonia, as a leader in e-government in the EU, announced in 2016 their partnership with the startup company Guardtime, in order to secure over 1 million patient healthcare records using a blockchain system. Their partnership with the blockchain startup company demonstrates that the emerging technologies, like blockchain in this example, can be used to protect sensitive healthcare records. The existing healthcare protocol requirements, such as the Health Information Exchange (HIE) and Integrating the Healthcare Enterprise (IHE), can be met using blockchain technology as a new form of data standardization for healthcare data distribution. Health and government organizations spend large amount of time and money in setting up and managing information systems and data exchanges. Blockchain's open-source technology, properties, and distributed nature can help reduce the cost of these operations. The blockchain-based Electronic Health Records might enable sharing and access to data, while securing it completely.

3.4. Relevant standards and legal frameworks

Long-term preservation of digitally signed records requires that they maintain their basic characteristics – authenticity, reliability, integrity and usability. In order to achieve that, it is necessary to rely on relevant standards and, usually national, legal framework.

3.4.1. ISO 15489 – Information and documentation – Records management

In 2001 the International Standards Organization (ISO) has published two documents, constituting two parts of a new international records management standard – ISO 15489 – Information and documentation – Records management. The first part relates to the standard in general, while the second provides technical guidelines previously known as a technical report. Through application of the ISO 15489-1:2016 – Information and documentation – Records management, Part one – Concepts and Principles, the ISO 15489-1:2001 was rendered obsolete.

According to the ISO 15489, e-records are required to maintain their authenticity, meaning that after the completion of every digital preservation procedure they are still required to be authentic, complete and useable, and must still keep the content, structure and context in relation to other preserved records.[17]

3.4.2. ISO 14721 – Open Archival Information System Reference Model

The Open Archival Information System Reference Model (OAIS RM) is one of the possible technical solutions for long-term preservation. The model was developed by Consultative Committee for Space Data Systems (CDSDS) with NASA in 1999. The model became an ISO standard in March 2003 (ISO 14721:2003). The latest version is from 2012 (ISO 14721:2012)[18] and is currently under review.

The OAIS reference model is applicable in any digital archive, but must be expanded and adapted according to the specific requirements of document creators and digital archives. The OAIS model refers to long-term digital preservation of records as objects and information packages (SIP, AIP, DIP). The model defines information model and functional model of a digital archive.

"ISO 14721:2012

- provides a framework for the understanding and increased awareness of archival concepts needed for long term digital information preservation and access,
- provides the concepts needed by non-archival organizations to be effective participants in the preservation process,
- provides a framework, including terminology and concepts, for describing and comparing architectures and operations of existing and future archives,
- provides a framework for describing and comparing different Long Term Preservation strategies and techniques,
- provides a basis for comparing the data models of digital information preserved by archives and for discussing how data models and the underlying information may change over time,
- provides a framework that may be expanded by other efforts to cover long term preservation of information that is not in digital form (e.g. physical media and physical samples),
- expands consensus on the elements and processes for long term digital information preservation and access, and promotes a larger market which vendors can support, and
- guides the identification and production of OAIS-related standards."[19]

3.4.3. DSS – Digital Signature Standard

The Digital Signature Standard (DSS) was issued in July 2013. This standard is a part of the official series of publications and standards relating to standards and guidelines adopted and promulgated under the provisions of the Federal Information Security Management Act (FISMA) of 2002. The name of these official series is "The Federal Information Processing Standards Publication Series of the National Institute of Standards and Technology (NIST)"[20]. This standard defines methods for digital signature generation that can be used for the protection of binary data (commonly called a message), and for the verification and validation of those digital signatures. Three methods are approved: 1) Digital Signature Algorithm (DSA), 2) RSA (Rivest-Shamir-Adleman) digital signature,

[17] ISO 15489-1:2016 (International Organization for Standardization, 2016)
[18] ISO 14721:2012 (International Organization for Standardization, 2012)
[19] Ibid.
[20] The Data Encryption Standard (DES) used to be part of these official series. It was in effect from July 1977 until May 2005. The algorithms described in this standard specified both enciphering and deciphering operations which are based on a binary number called a key. (National Institute of Standards and Technology, 1999)

and 3) Elliptic Curve Digital Signature Algorithm (ECDSA). Except methods, this standard includes requirements for obtaining the assurances necessary for valid digital signatures. Methods for obtaining these assurances are provided in NIST Special Publication (SP) 800-89, Recommendation for Obtaining Assurances for Digital Signature Applications. (National Institute of Standards and Technology, 2012)

3.4.4. eIDAS Regulation

The eIDAS Regulation, fully titled "Regulation (EU) no 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC", applies in all EU Member States since 1 July 2016. The transition period ended on 1 July 2017 and the Regulation is fully implemented and the trust services are no longer regulated by national laws. The eIDAS Regulation has set rules under which persons and institutions can use the means of electronic identification provided by their own states in other member states, rules for various trust services, especially electronic transactions and a legal framework for "electronic signatures, electronic seals, electronic timestamps, electronic documents, electronic delivery services and web page certification services". Since the eIDAS regulation requires, among other things, that the digital signature certificates are issued only to natural persons, all digital signature certificates issued to legal persons, under the superseded eSignature Directive, cannot be used to create legally valid digital signatures any more. Legal persons should be issued signing certificates to be used as digital seals. Further, the eIDAS requires establishment of a list of qualified trust service providers in European Union (the EU Trusted list). Concerning long-term preservation of digitally signed records, the eIDAS Regulation defines e-timestamps as "information on an electronic media which creates a bond between other electronic media data with a certain time, thus proving their existence at a specific time and date". This regulation is expected to have a positive effect on businesses and persons by enabling the creation of new qualified trust services.[21]

"The eIDAS Regulation sets rules for the preservation of electronic signatures, electronic seals or certificates related to trust services. Preservation is different from electronic archiving (which is NOT a trust service under eIDAS). The objectives and targets of the process will make a distinction between the two activities:

- Preservation under eIDAS aims at guaranteeing the trustworthiness of a qualified electronic signature or qualified electronic seal through time. The technology underpinning such trust service therefore targets the electronic signature or seal;
- Electronic archiving aims at ensuring that a document is stored in order to guarantee its integrity (and other legal features). The technology underpinning electronic archiving therefore targets the document. Electronic archiving remains the competence of Member States.

In other words, electronic archiving of documents and preservation of electronic signatures and electronic seals are different in nature, are based on different technical solutions (attached to the document or to the electronic signature/electronic seal) and differ in their finality (conservation of the document vs. preservation of electronic signature/electronic seal)." (European Comission, 2016) Interestingly, the EC uses the term *conservation* of the document and differentiates it from the term *preservation* of the e-signature while this research team thinks that in both places the term *preservation* should be used.

3.4.5. ISO/TC 307 – Blockchain and Distributed Ledger Technologies

Blockchain is in the process of standardization by the International Standardization Organization (ISO/TC 307)[22] with the aim to support interoperability and data interchange among users, applications and systems. The 1st meeting was held in April 2017. Members of this study group are actively involved in the development of the standard – Victoria Lemieux is appointed Head of the Terminology Working group and Hrvoje Stančić is appointed President of the Croatian ISO/TC 307 Mirror Technical Committee with the Croatian Standards Institute.

[21] eIDAS Regulation (European Parliament, 2014)
[22] ISO / TC 307, https://www.iso.org/committee/6266604.html

In December 2017 CEN/CENELEC created a Focus Group on Blockchain and Distributed Ledger Technologies with the aim to identify specific European standardization needs, to map these needs (including blockchain and DLT governance in the frame of the General Data Protection Regulation – GDPR), with the current work items in ISO/TC 307 and to encourage further European participation in this ISO Technical Committee[23].

3.5. Current approaches to long-term archiving and preservation of digitally signed records

Long-term preservation of digital records that are digitally signed or have a digital seal attached to them is a challenge for the archival profession. Such digital records are not easy to preserve, not only because of the constant technology advancements, but also because the certificates they rely on are not designed with unlimited duration in mind..For example, the Financial Agency (FINA), a CA in Croatia, issues certificates with the validity period of two years and the Agency for Commercial Activities (hrv. Agencija za komercijalnu djelatnost, AKD) with the validity period of five years (those are used in e-Identity Cards). The root certificates of the issuer generally have a longer validity period, e.g. ten years. After the expiration of the validity period of the certificate, it will no longer possible to check the validity of the digital signature using common methods and software, but it will still be possible to check the integrity of the record itself. Currently there are several approaches to long-term preservation of digital records that have digital signatures or seals attached to them.

According to PREMIS (Data Dictionary for Preservation Metadata: PREMIS version 3.0, 2015), preservation repositories use digital signatures in three main ways:

1. *For submission to the repository*, an Agent (author or submitter) might sign an object to assert that it truly is the author or submitter;
2. *For dissemination from the repository*, the repository may sign an object to assert that it truly is the source of the dissemination;
3. *For archival storage*, a repository may want to archive signed objects so that it will be possible to confirm the origin and integrity of the data.

In all cases where digital signatures are used by the repository as a tool to confirm the authenticity of its stored digital objects over time, must the signature itself and the information needed to validate the signature be preserved (along with originals and supporting documentation – digital certificates, CRLs, OCSP responses etc.). Revalidation of the signatures should be avoided whenever possible until archiving technology is fully established and supported by legislation and legal practice.

According to Blanchette (Blanchette, 2006), from the point of view of archives there are three possible options:

1. *Preserve the digital signatures:* This solution supposes the deployment of considerable means to preserve the necessary mechanisms for validating the signatures, and does not address the need to simultaneously preserve the intelligibility of documents;
2. *Eliminate the signatures:* This option requires the least adaptation from archival institution, but impoverishes the description of the document, as it eliminates the signature as one technical element used to ensure the authenticity of the documents[24];
3. *Record the trace of the signatures as metadata:* This solution requires little technical means, and records both the existence of the signature and the result of its verification. However, digital signatures lose their special status as the primary form of evidence from which to infer the authenticity of the document. Moreover, this approach requires the existence of a trusted third party to preserve and authenticate the metadata.

Other possible approaches are e.g. using official state registers of created/received records in combination with early archiving. Certain authors argue that the only option is the first one, i.e. to

[23] CEN and CENELEC's new Focus Group on Blockchain and Distributed Ledger Technologies (DLT), https://www.cencenelec.eu/news/articles/Pages/AR-2017-012.aspx

[24] It is widely considered bad practice to discard signature element altogether. One will hardly be able to convince a court that this was done in good faith. Elimination is more often understood now as a refusal of an archival institution to do re-validation.

develop a Trusted Archival Service (TAS) which could guarantee that the signature of a record can still be validated years later (Dumortier & Van den Eynde) but this approach still has to see its widespread implementation.

However, results of the previous InterPARES projects recommend the third option, i.e. to organize a digital archive in the way to check the validity of the digital signatures at the ingest phase (either by technical re-validation of signatures, or by obtaining assurances from the relevant authority), add the validity information to the records' metadata, and preserve the records without addressing the digital signature's validity further. Thus, the issue of trust is shifted from the (digitally signed) record to the archive preserving digital records and the associated (validity) metadata. This follows the more traditional model of archival preservation, which stands in contrast to the underlying premise of blockchain and distributed ledgers technology as not reliant upon a trusted third party or preservation intermediary. (Nakamoto, 2008)

Later in this research report we will show that there is a fourth option which is based on the principles of blockchain and distributed ledger technologies, i.e. to register the validity of the digital signature in a blockchain.

3.5.1. OAIS and TDR

The digital archive established according to the OAIS reference model is considered the least technically demanding solution. By entering a record with a digital signature in a digital archive, the validity of the digital signature might be verified and that information might be recorded in the metadata. After verification the record is stored in an archival information package (AIP) with the associated metadata. In this way, if the validity information indeed was recorded as metadata, the expiration of the certificate is no longer so important because the information about its validity at the time of ingest was stored. However, to have confidence in the OAIS-compliant digital archive it should be established in accordance with the ISO 16363:2012 Audit and certification of trustworthy digital repositories (TDR) (International Organization for Standardization, 2012), which prescribes how numerous steps taken during long-term preservation should be carried out in the way not to endanger the credibility of digital records stored in the archive. In other words, only when a digital archive is made OAIS and TDR compliant, it can be trusted enough to transfer the information about the validity of a digital signature or seal from the record itself to the digital archive.

3.5.2. CRL and OCSP

In the context of technologies that contribute to the long-term preservation of digitally signed records, it is necessary to distinguish between Certificate Revocation List – CRL (Cooper, Santesson, Farrell, Boeyen, Housley, & Polk, 2008) and the Online Certificate Status Protocol – OCSP (Santesson, Myers, Ankney, Malpani, Galperin, & Adams, 2013). Adding information about the validity status of the certificate to a CRL or OCSP response, and by including a certification chain for the sake of trust in the digitally signed certificate, it is possible to successfully validate a digital signature after the expiration of the signature certificate.

The CRL list is the usual way to revoke a certificate. The CRL list itself represents a standard file with a series of serial numbers. Each digital certificate has a unique number, so if it is found on the list it means that the digital certificate was revoked. The CRL is published by the appropriate Certification Authority (CA) at predetermined time intervals. However, this is a disadvantage of CRLs because they do not function as a real-time service On the other hand, once obtained the CRL may be used without establishing an online connection until CA issues new version. The mandatory fields crucial for later digital signature verification are serial number, date, and time of the digital certificate revocation.

The OCSP service is based on the OCSP protocol, developed due to the need to overcome the CRL-related shortcomings. The OCSP in the PKI infrastructure is the responsibility of Validation Authority (VA), which validates digital certificates. In the case OCSP is used, the CA sends the information about the digital certificate revocation to the VA. A person or service wishing to verify the validity of a digital certificate sends a query to the VA and receives a "valid", "recall" or "unknown" response. Therefore, by using OCSP it is not necessary to download CRL and save it along with the archived records, but only to establish a direct communication with the VA to be able to check if a digital

certificate is valid. At the same time this is the main disadvantage – the loss of Internet connection disables the validity checking.

3.5.3. Archival Timestamp

Besides the standard timestamp, there is also a special type of timestamp intended for long-term preservation – the archival timestamp or *timestamp token for long term availability and integrity of the validation material*. It differs from a standard timestamp neither theoretically nor technically, but only by its scope. It includes a number of hash values – one of every piece of information that needs to be kept and bound together for a long time. The primary purpose of the archival timestamp is that its application extends the validity of the digital signature and enables the positive CRL or OCSP response to the validity check even after the validity period of signature certificates.

Archival timestamp implements an onion-like wrapping principle. The standard ETSI EN 319 102-1 Procedures for Creation and Validation of AdES Digital Signatures; Part 1: Creation and Validation defines four basic levels of baseline digital signatures enabling interoperability and life-cycle of records. Each level wraps all previous information. The levels are: 1) B-B – basic, 2) B-T – timestamp added to the B-level, 3) B-LT – long-term validity verification information are added to the T-level, and 4) B-LTA – enables periodic addition of the archival timestamp to the LT-level (Figure 1). (ETSI, 2016) In practice, this means that the long-term preservation system should be set up in a way that it checks the expiration period of certificates of every digitally signed record or the expiration of already added archival timestamp and (re)applies archival timestamp (B-LTA) before the signing certificates or previously added archival timestamp expires. This, of course, may present a challenge for the digital archives preserving large collections of digitally signed records.

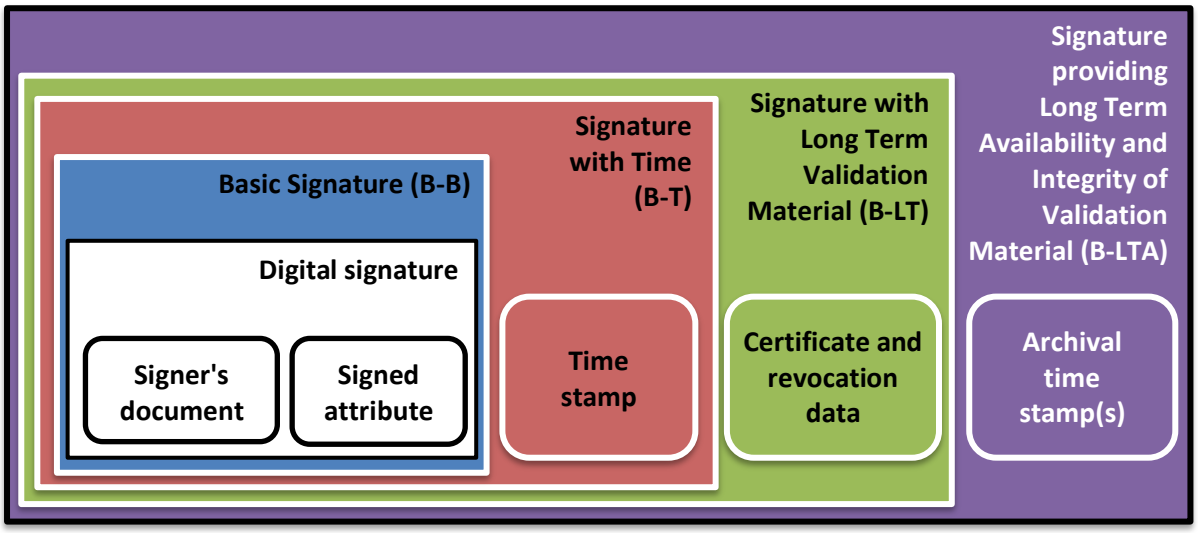


Figure 10. Archival Timestamp

4. Research questions

The questions raised by the research team at the beginning of the research where:

1. Why the digitally signed, timestamped or sealed records should be preserved?
2. Are they still of any business or historical value that justifies the expenses?
3. Is there a need to revalidate already expired digital certificates and how it can be done?
4. Are there any official databases containing the information from the records?
5. Can digital certificates be renewed by timestamping the encapsulated certificates prior to their expiration?
6. Are there other ways to keep the signatures valid for the long-term?
7. In what situations it is important to require a timestamp?
8. Can we presume trustworthiness of the time of creation if it is recorded outside of the controlled environment of the digital archive?
9. Can we presume trustworthiness of digital signatures which do not have revocation information embedded, or can we presume trustworthiness of the ingest server's digital signature if it does not have a timestamp?
10. Are there any existing legal ways to authenticate the body of records without revalidating original signatures?
11. Is it possible to do nothing or use not exactly 100% legally-compliant solutions? What might be the consequences?
12. Is there a need to amend the laws and regulations?
13. Using contemporary technology, is it possible to develop systems which would provide trust in digital signatures and certificates for longer periods?

5. Aims and Goals

The aims of the research study:

1. To addresses an important archival issue of preservation of digital records in the cloud by utilizing new technological concepts, i.e. blockchain.
2. To build a model suggesting how to preserve the trustworthiness of the digital records with digital signatures, certificates, timestamps or seals added to them.
3. To investigate the possibilities of revalidation of the expired digital signatures, periodical re-signing of digital records or renewal of timestamps, addition of archival timestamps, injection of additional (timestamped) proof of existence etc.

The goals of the research study:

1. To achieve results that could be used to draft and/or to improve regulatory frameworks.
2. To achieve results that could be used to draft and/or to improve internal organizational policies and procedures.
3. Overall, to achieve results that are relevant for organization and development of trusted archival services relying on ingest of trustworthy records.

6. Methodology

To achieve the stated aims and goals, the research was divided into five phases over 19 months (March 2016 – September 2017). In phase one of research the team used a comparative approach to review existing literature in the field of archiving, legal framework, blockchain and distributed ledger technologies. In the second phase three cases were developed. They investigated the procedures and challenges in real-life cases of long-term preservation of digitally signed records in public services. In the third phase, which was overlapping with the second phase, various recordkeeping and archival preservation use cases were investigated in the course of the development of case studies and based on that, in the fourth phase, the model for preservation of validity of digitally signed and timestamped records was developed (TRUSTER VIP (Validity Information Preservation) Solution: TrustChain). The model provides a framework for a system which could enable archives to preserve authenticity (certificate information or, at least, certificate validity information) and integrity (using digital signatures) of digitally signed documents. The fifth, and final, phase of the research consisted of writing this report (Figure 11).

InterPARES Trust

Model for Preservation of Trustworthiness of the Digitally Signed, Timestamped and/or Sealed Digital Records (TRUSTER Preservation Model) (EU31)

	2016										2017								
	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19
Phase 1 - review of the relevant literature																			
	FHSS																		
	IE																		
	UBC																		
Phase 2 - development of case studies																			
				T															
				F															
				E															
Phase 3 - testing of various recordkeeping and archival preservation use cases																			
				FHSS															
				E															
				UBC															
Phase 4 - development of the model for preservation of validity of digitally signed and timestamped records																			
									FHSS										
									T										
									F										
									E										
									IE										
									UBC										
Phase 5 - write a report																			
																FHSS			
																T			
																F			
																E			
																IE			
																UBC			
Conference call meetings (using Victoria Lemieux's access to "GoToMeeting")																			
Conference calls		X		X			X		X		X		X		X	X		X	X
Milestones		M1		M2		M3			M4			M5		M6		M7		M8	
Result			R1				R2		R3						R4				FR
Meetings and dissemination																			
EU Team meetings (Marburg, Oslo, Girona)				8.-10.				24.-26.							11.-12.				
Plenary meeting (Rio)												20.-22.							

M	Milestone
R	Result (FR = Final Result)
FHSS	FHSS
T	Teched
F	Fina
E	Enigio
IE	Independent Expert

Figure 11. Research timeline

7. Findings

The findings are divided in two sections – the first one outlining the results of the case studies, which are available as stand-alone products, and the second one being the TrustChain information system model for long-term preservation of digitally signed records.

7.1. Case studies

In cooperation with three involved partners, FINA (Croatia) TechEd (Croatia) and Enigio Time (Sweden) three case studies have been developed. Each of these three partners had access to digitally signed records with expired or soon expiring digital signatures. The conducted case studies have explored how various financial, public sector and medical institutions deal with long-term preservation of digitally signed records.

In the course of research, the questionnaire was developed and used in order to make the case study results comparable. It was used in the case studies 1 and 2. The questionnaire can be found in Appendix 1 of this report.

The main goals of the case studies were (1) to analyse the current use of and state of preservation of digitally signed records held by different institutions, (2) to understand the perceived value of the need for archiving of the digital signatures as well as the archiving of the validity of the digital signatures, and (3) to understand how could the expiration of certificates in digital signatures influence the admissibility of records as evidence in the court.

Overall, the case study analysis was focused on the digitally signed records and the preservation of the validity of used certificates in various recordkeeping and archival preservation use cases. Thus the case studies summarise the procedures with those records related to the case study goals. They could also function as a foundation for further cooperation or additional, more detailed studies.
The full texts of the case studies are available as stand-alone products. Here, only brief information on the case studies is given.

7.1.1. Case study 1 – digitally signed retirement fund records

This case study has been conducted in cooperation with FINA – Financial Agency in Croatia between June and October 2016.

The case study investigated the processes with digitally signed records in a defunct e-Regos system which was transferred to FINA as an outsourcing service. The research showed that the records in the e-Regos system used FINA's (a CA in Croatia) qualified signatures and timestamps. Before the e-Regos system was discontinued, the records were transferred to the FINA who now stores the original digital records locally, i.e. they are not accessible online. However, the information from the records was transferred to an online database and it is accessible there. The digital signatures on the original records cannot be validated any longer because the CRLs and certificate chains from that time are not preserved.

The study highlights the need for development of a digital preservation strategy and a policy for preservation of the validity of digital signatures. It was also noted that in this specific case an organizational solution is possible – legal preservation obligation may be shifted to a database, and original digitally signed records may be disposed of, since they have no lasting legal or business value. This approach is straightforward and requires no technical innovations, but before the disposal of original documents updating legislation and regulations will be needed and may prove to be difficult.

7.1.2. Case study 2 – digitally signed e-tax records

This case study has been conducted in cooperation with TechEd Consulting Ltd. and Croatian Tax Administration, between January and April 2017.

The case study investigated the processes with digitally signed records in custody of the Croatian Tax Administration. The research showed that the records have XML signatures, qualified signatures and timestamps. The digital signatures on the original records dating from 2006 cannot be validated any

longer because they have expired. In order to check them the Tax Administration should request the CA which originally issued keys for digital signatures to provide the old CRLs and certificate chains from that period. However, they are not preserved. Neither the information on the validity of digital signatures nor other information from or about digital signatures were recorded as metadata[25].

7.1.3. Case study 3 – digitally signed medical records, procurement and supplier contracts, official political decisions and minutes of meetings

This case study has been conducted in cooperation with Enigio Time and Region Skåne (Sweden) between June and November 2016.

This case study focused on the most important and the largest volume of records using digital signatures in the Region Skåne – the digitally signed medical records, procurement and supplier contracts, official political decisions and minutes of meetings. Medical records are the biggest record type created and handled by Region Skåne. It constitutes about 80% of the records stored in the Region Archive.

Digital signatures of many different kinds are used in many different ways in all types of records and systems. Value of preservation was not fully recognised or clearly stated in the workflows. When a digitally signed record is archived the validity of the digital signature is not verified. The information from the digital signature is saved as metadata. The records are kept in the source system just in case a signature is needed for verification. No other expressed strategy or a proprietary process being able to recover or prove signature validity was noted as existing at the Region Skåne. The study triggered discussions about the need for a more focused analysis and development of a common strategy for digital signatures' validity preservation.

7.2. TRUSTER VIP (Validity Information Preservation) Solution: TrustChain

7.2.1. Introduction

The TRUSTER VIP Solution TrustChain is the fourth approach for long-term preservation of digital records that have digital signatures or seals attached to them (c.f. chapter 3.5. of this report). The model behind the TrustChain allows archives and other institutions dealing with digitally signed or sealed records to avoid having to re-sign (or timestamp) records periodically, before their digital certificates expire.

To briefly sum up what was explained earlier, digital signatures rely on the public key infrastructure (PKI) concept which enables users to create a hash of a document (a signature) using private (and secret) key. Once this is done anybody can confirm that the document has indeed been signed using that private key by recalculating the hash using a public key connected to the private key (and its owner in the case of advanced digital signatures). The purpose of a digital signature is twofold. Firstly, it guarantees data integrity, i.e. a digital signature can be used to confirm a document has not been tampered with after it has been digitally signed. Secondly, it can be used to identify the person or institution (in case of digital seals) which signed it. This information (the connection between a signatory and an actual identity) is stored in the digital certificate which is issued by a certificate authority (CA). These certificates have a limited lifetime. Once a certificate reaches the end of its lifetime it can no longer be used to identify the signer by common methods and software and, depending on the requirements of the document or the institution archiving it, might need to be re-signed or timestamped. Certificate expiration is a necessity due to the evolving security standards, development of the information technology which, in time, weakens the key strength, and the possibility that the keys become compromised.

[25] The authenticity of the whole set of records can be established by a court in the basis of available circumstantial evidence. This is usual practice in common law countries which, unfortunately, is not widely adopted in continental law countries.

The results of the case studies have demonstrated a great need for a standardized system for long-term recordkeeping or archiving of digitally signed or sealed records. While industry standards address this problem by relying on the timestamping services[26] their solution suffers from a problem similar to the digital signatures themselves. Timestamps can be described as digital signatures which only guarantee data integrity from a time perspective. Most timestamping services are linking-based schemes or PKI based solutions. A PKI based "timestamp" includes a hash of the timestamped document, time of stamping and is, most often, signed with the private key of the timestamping service. Therefore, it is evident that this system suffers from the same technological lifetime limitation as the digital signature itself, although it is usually much longer (5-20 years). The timestamped document will need to be stamped again after a certain period – just before the certificate used by the timestamping service expires, the cryptographic algorithm used becomes obsolete or before their private key becomes compromised (of course, highly uncertain if and when). It should also be noted that some timestamping services use a public key without a certificate under a transient-key scheme[27]. While absence of a certificate and security inherent in temporary keys makes timestamps appear to have even longer lifetime (possibly limitless), such approaches still does not support certificate preservation and rely on security of the persistent key list. On the other hand, the linking-based schemes are somewhat similar to the TrustChain model but generally do not check certificate validity and thus only guarantee data integrity.

The concept of the TrustChain has been published in the INFuture2017 conference paper "A model for long-term preservation of digital signature validity: TrustChain" (Bralić, Kuleš, & Stančić, 2017). However, the model is still in an early, conceptual, phase and is going to be developed further. Therefore, the model to be found in this report is an evolution from the model presented in the original paper. The basic concept of the model is the same, but it is refined at the level of technical details.[28]

In the next phase, we expect further refinement of the model and development of working prototypes of the system. Once this has been achieved we can move on to recognizing and recruiting institutions interested in using the TrustChain system and maintaining the infrastructure it requires.

7.2.2. The TrustChain model

The goal of the TrustChain is to enable archival institutions (or others with such needs) to avoid having to periodically re-sign (or timestamp) all their archived, digitally signed records. We envision TrustChain, a blockchain-based solution, being maintained by an international alliance of archival institutions. The system could also be implemented by a single institution but then the degree of security in signature validity is significantly reduced.

[26] ISO/IEC 18014-3 (International Organization for Standardization, 2009) and ETSI 319 422 (ETSI, 2016).
[27] As described in the ANSI ASC x9.95 standard (American National Standards Institute, 2016).
[28] The research is also part of the Vladimir Bralić's PhD thesis.

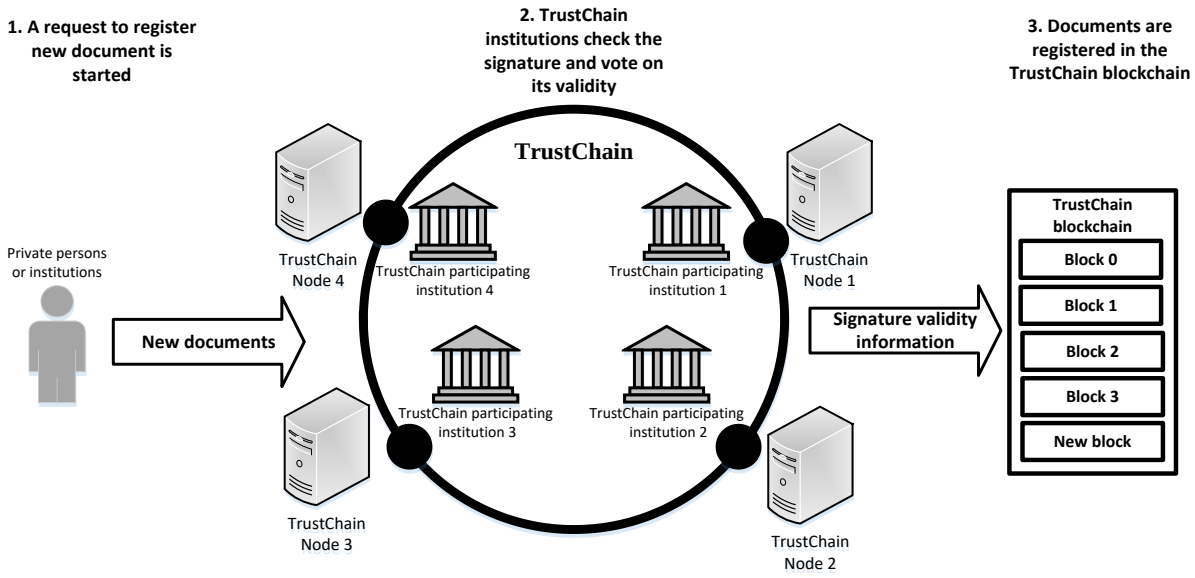


Figure 12. TrustChain concept

The TrustChain achieves the stated goal by checking document's signature validity and, if valid, writing the signature's hash (and possibly some document metadata) in the blockchain. Signature validity is checked by all or, if their number is sufficiently high, some of the participating institutions. If the signature is deemed valid the information is permanently stored in the TrustChain blockchain (Figure 12). The process of registration of a document in the TrustChain is shown in Figure 13 (Bralić, Kuleš, & Stančić, 2017).

The blockchain itself would probably be a publicly available ledger and anybody might be able to read it. However, if the participating institutions require it, the blockchain could certainly be a permissioned ledger available to authorized nodes only. Registering new data, documents or records in the blockchain could also be a publicly available option, or limited only to the participating institutions or possibly commercially available with a processing fee being charged by the TrustChain nodes.

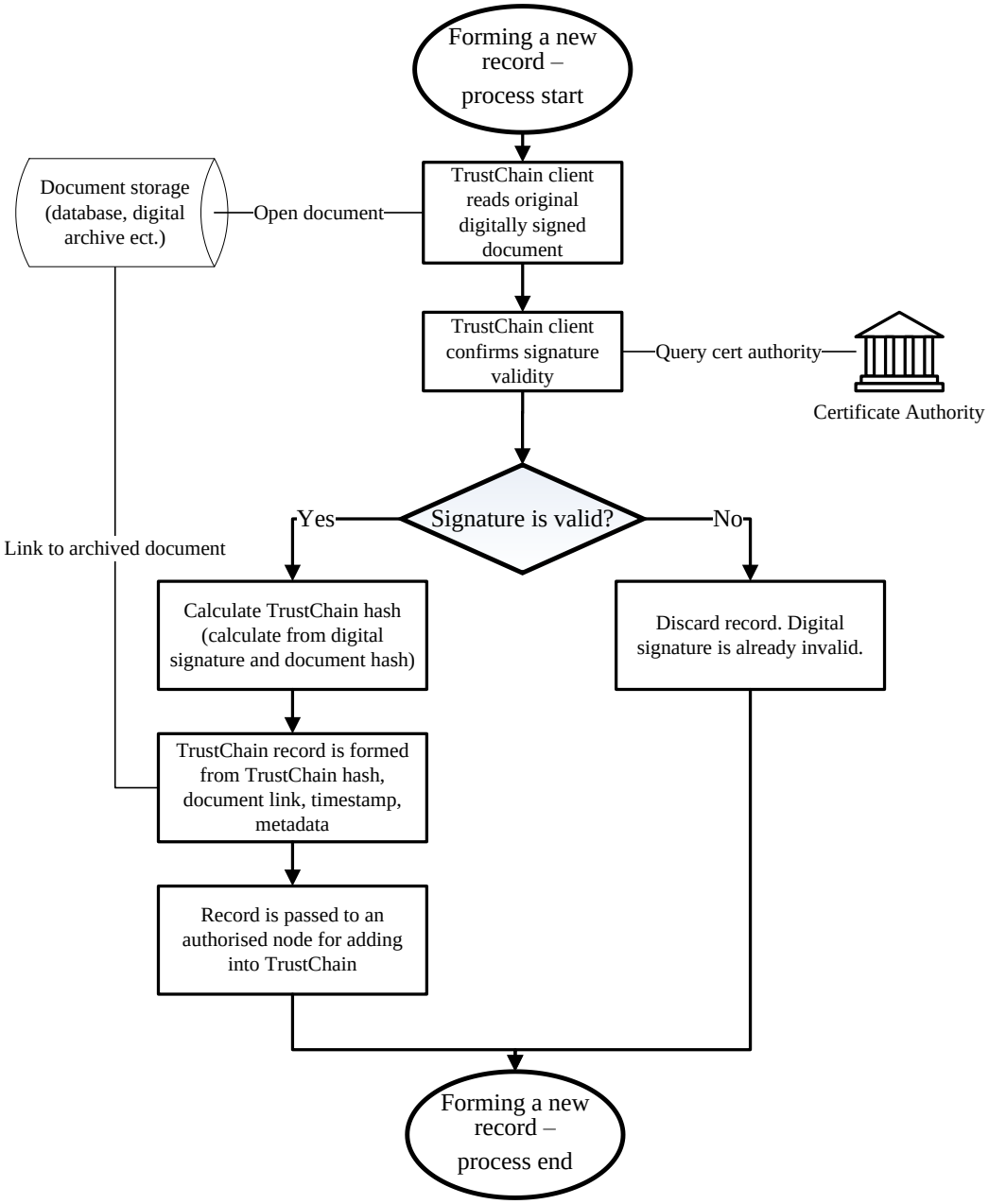


Figure 13. Registration of a document in the TrustChain

Having signature validity information written into an immutable blockchain provides evidence that the signature was valid at the point of its TrustChain record creation and that neither the record (as assured by the signature) nor the blockchain entry has been tampered with since then. This structure is illustrated in Figure 14. Because each block contains hash of the previous block it is impossible to change a fragment of a previous block without rewriting all the subsequent blocks. Further, previous block hash is written as part of the voting information, which means the voting information is also protected by hash. The votes are formed with a private key belonging to nodes voting on a block.

An attacker looking to alter information written into the blockchain would need to change all subsequent blocks and to achieve this (s)he would need access to private keys of all member institutions which have voted on blocks after the one being changed. Depending on the number of participating institutions this might be a practical impossibility. That is why the TrustChain is considered to be more secure with more participating institutions. Not only that the increase in the number of the participating institutions means that more private keys would need to be compromised, but also the diversity of the institutions taking part in the voting process ensures more confidence that the signature indeed was valid at the time of its addition to TrustChain.

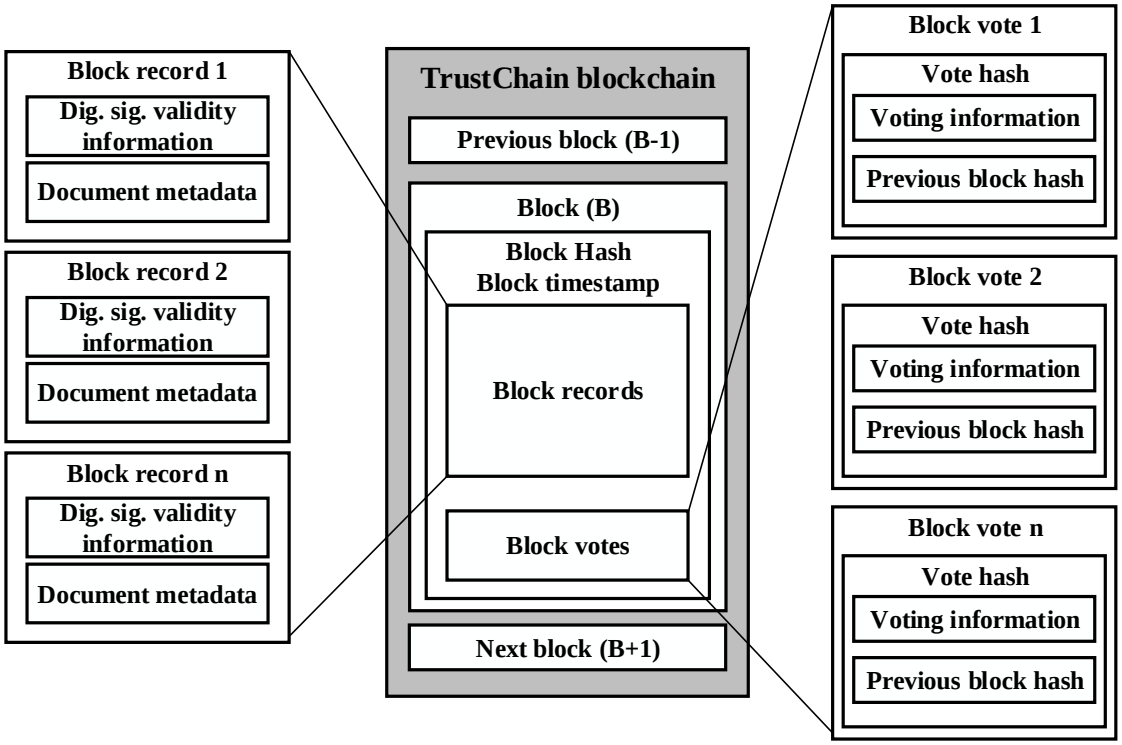


Figure 14. TrustChain blockchain structure

While it is true that like the alternative methods (timestamping) TrustChain relies heavily on the PKI concept, and thus on the asymmetric cryptographic algorithms, there is no need to periodically re-sign the data because of the way the data is stored. TrustChain's blockchain data structure is what enables it to avoid the pitfalls of the compromised private keys or outdated cryptographic algorithms. Once an algorithm becomes outdated TrustChain voting nodes merely need to change the algorithm they use to sign their votes. Blocks signed with a compromised or outdated key do not need to be re-signed. If an attacker wanted to re-sign them with altered data (s)he would, again, need to change and re-sign all the subsequent blocks as well and eventually (s)he would run into a block signed by an uncompromised key (and a safe algorithm). In short we could say that (as in other linked data models) adding a new block to the TrustChain re-signs (or re-confirms) all the existing entries or that the entire (block) chain is as strong as its strongest link, or in our case the block which will always be the most recent one (Figure 15).

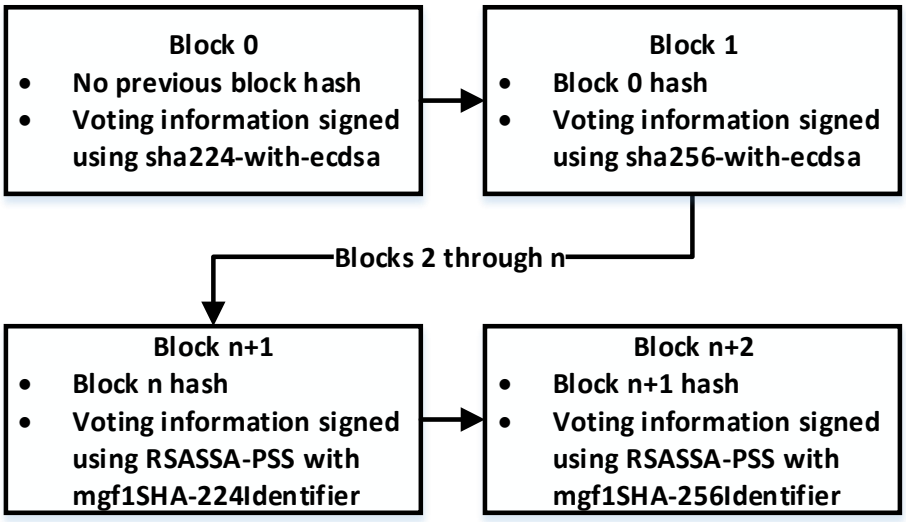


Figure 15. TrustChain voting algorithm changes

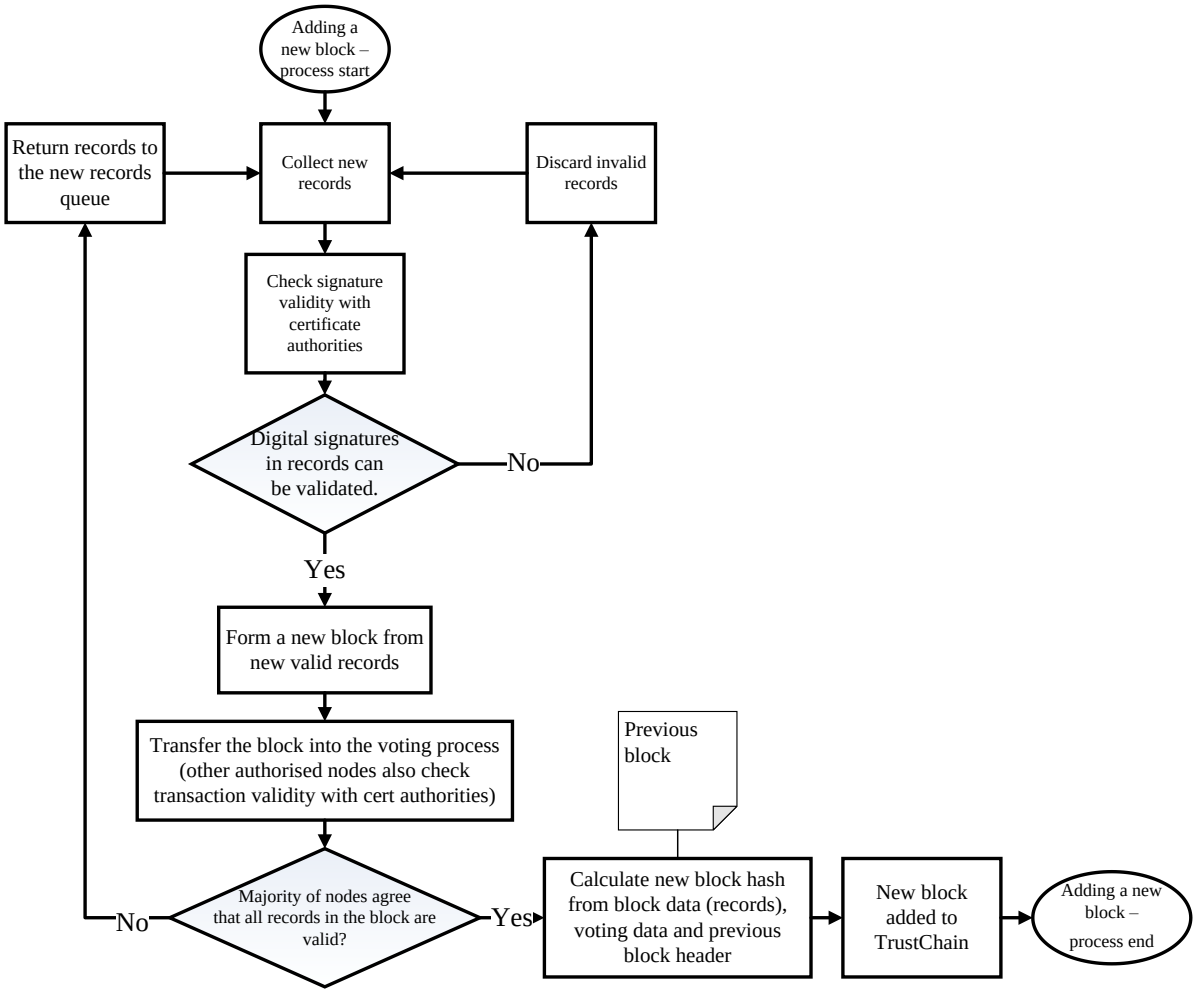


Figure 16. Adding a new block to the TrustChain

The process of adding records to a block and writing that block into the blockchain is left exclusively to TrustChain nodes. The nodes collect new (candidate) records from a queue and attempt to validate all signatures. If a signature fails, the record is discarded as invalid and new records are collected. Once a sufficient amount of valid records is found, they are added to a block but only after other nodes confirm signatures' validity of the records as well. The required number depends on the total number of available TrustChain nodes and the required level of reliability (the more nodes rechecking the records, the more reliable the vote will be). Since the number of participating institutions is not known, at this, early stage we will assume that all participating institutions maintain a node and they all vote on every block. Should the number of institutions rise to a number where having everyone vote becomes a performance issue, a smaller, randomly selected, subset of nodes can vote for each block. This subset should change for every block. If the majority of voting nodes agree that the block is valid it can be added to the blockchain (after having its hash calculated from its contents and the previous block's hash). Otherwise, the block is discarded and the records that formed it are returned to the new records queue (Figure 16). (Bralić, Kuleš, & Stančić, 2017)

The confirmation process of the (expired) digital signatures begins with finding the relevant records in the TrustChain blockchain. For this, the TrustChain relies on the recorded document metadata – ISAD(G)'s essential set of elements but also may contain information pertaining to the archival bond (Lemieux & Sporny, 2017). Once the relevant record is identified, all that needs to be done is to recalculate the hash from the original document and compare it to the one written in the TrustChain. If these hashes match, one can reliably claim that the document and its signature have remained unchanged since the date indicated by the blockchain record timestamp (Figure 17). (Bralić, Kuleš, & Stančić, 2017)

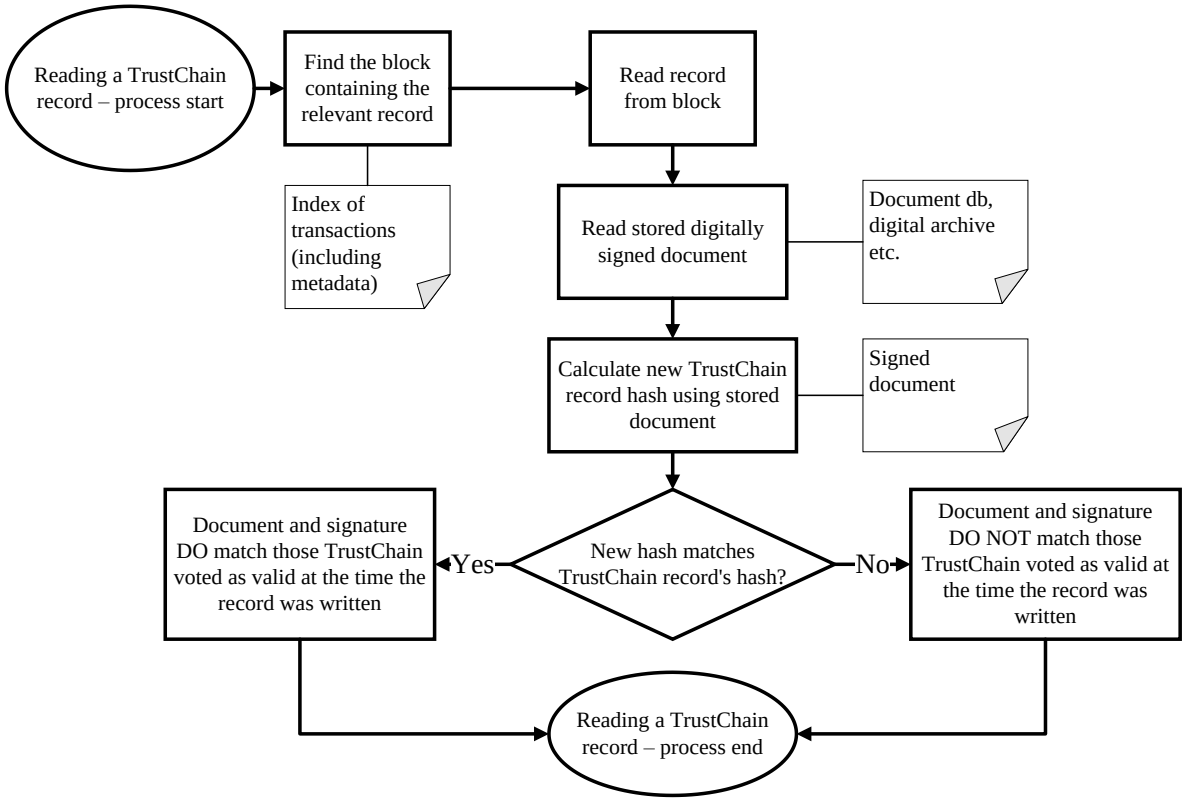


Figure 17. Reading a record from the TrustChain

7.3. Discussion

The current model of the TrustChain detailed here has been discussed by the InterPARES Trust researchers (those not directly taking part in the development of the model) and by the INFuture2017 conference[29] audience and reviewers. Several points which have been raised will be addressed next.

Firstly, there are opposing views regarding the inclusion of metadata in the blockchain. On the one hand, it is convenient to have some of this information written directly into the blockchain to enable searching of the blockchain. Without it searching would only be possible by the hashes, i.e. only those in possession of the document's hash would be able to identify it within the chain while others would not. Writing metadata in the blockchain supports openness of the system and enables anyone to confirm its contents. On the other hand, not writing metadata encourages privacy on the blockchain, which might be important for certain users, for example banks or other financial institutions. One of the possible solutions to this is to create different types of metadata according to ontologies specific for certain document types. This might even be a preferable solution as it would allow a lot of flexibility for the TrustChain users. One user could, for example, create a confidential type entry which might include nothing but a document's hash and timestamp while another user might use a complex ontology for his/her documents which would make them easily identifiable by anybody and suitable for indexing. The creation of such ontologies would likely mimic existing archival standards.

Secondly, some critics did not see the need for such a system or saw it as overcomplicating the existing timestamping schemes. The reason for the development of the TrustChain was to preserve certificate information (or at least preserve information about its validity at a certain point in time). To the knowledge of TRUSTER Team researchers none of the existing timestamping services meet this requirement. Of course, successful implementation of the TrustChain system requires significant infrastructure and willing participating institutions as well as further development while timestamping services are standardized and commercially available right now, but in the view of the team researchers the TrustChain is a better solution. The TRUSTER VIP Solution "TrustChain" demonstrates that such a system is possible and that periodical re-signing (re-timestamping) of records is not necessary.

8. Conclusions

A suggestion for enhancing TurstChain security is to add a third party qualified timestamp to each block. While this further complicates the system it is a good suggestion and would indeed make the blockchain safer. The current model did not, in detail, consider its own timestamps. At this point it was simply assumed that nodes creating blocks and votes would use their own internal server times. This is not a good solution and is certainly insufficient from a security standpoint because it leaves a clear, and quite possibly, easily exploitable attack vector. Reliance on timestamps, outside clocks and internal clocks is something which will be addressed in the future model development and research.

Another line of current research is implementation of a transient-key scheme in the TrustChain either as a built-in feature or relying on an outside service. A transient-key is a variation of the usual PKI scheme in which keys are issued to a very short period and are tied to a certain time interval instead of a person or an institution. A keys lifetime in such a system is usually limited to a few minutes and a list with all previous interval keys is published and maintained by several publicly available sources. Such a (third-party) timestamp could be added to each block to use as a trusted outside source of time and fulfil the requirements mentioned earlier. Alternatively, a transient-key scheme could be built into the TrustChain itself. A transient-key scheme is patented (Doyle, 2002), and that might prove to be a major stopping point for implementing such a system directly but such a system could still be used as an outside timestamp if the team decides such a solution is preferable to a link-based scheme.

A possible evolution of the TrustChain system, as suggested by Enigio Time – partners at this research, is to also store validity intervals of the certificates themselves (and their chains) instead of digital signature validity information. If the certificates' validity periods are stored in an immutable data structure such as a blockchain, one could also confirm that the documents which had been timestamped before their signature certificate expired had a valid certificate and thus signature at the

[29] INFuture2017: Integrating ICT in Society, Zagreb, 8-10 November 2017, http://infoz.ffzg.hr/infuture/

specified time. That could offer proof of their authenticity in addition to the data integrity insured by their (still valid) timestamp. In the long run, an official combination of such a TrustChain and the PKI infrastructure could possibly remove the problem of expiring certificates and benefit a well-established technology like PKI in combination with the new blockchain technology. A data structure and a voting system similar to the one already presented might be suitable for storing this information as well. The benefit of such a system is that already expired signatures can still be controlled if they have a valid timestamp from the time signed. At this time both variations are being further developed and are intended to function independently of each other. Working titles of the two variants are: TrustChain-H (preserves Hashes) – the original proposal detailed in this report, and TrustChain-C (preserves Certificates) to the new system proposal yet to be developed.

9. Products

Case study 1 – digitally signed retirement fund records
Case study 2 – digitally signed e-tax records
Case study 3 – digitally signed medical records, procurement and supplier contracts, official political decisions and minutes of meetings
Blockchain bibliography
Blockchain terminology – in the InterPARES Trust terminology database

10. List of figures and tables

List of figures

Figure 1. Example of hash values....10
Figure 2. Example of pseudorandom characteristic of hash function....11
Figure 3. Merkle tree....11
Figure 4. Three types of network topology....12
Figure 5. Linking of the hash values....12
Figure 6. Blockchain creation....13
Figure 7. Multiple hashes combined in one block....13
Figure 8. Hash modification propagation through the blockchain....14
Figure 9. Checking of a hash value in the blockchain....14
Figure 10. Archival Timestamp....23
Figure 11. Research timeline....26
Figure 12. TrustChain concept....30
Figure 13. Registration of a document in the TrustChain....31
Figure 14. TrustChain blockchain structure....32
Figure 15. TrustChain voting algorithm changes....33
Figure 16. Adding a new block to the TrustChain....33
Figure 17. Reading a record from the TrustChain....34

List of tables

Table 1. The differences between public and private blockchain....15

11. References

American National Standards Institute. (2016). Retrieved from ANSI X9.95-2016 Financial Services - Trusted Time Stamp Management And Security: https://infostore.saiglobal.com/en-gb/Standards/ANSI-X9-95-2016-1894464/

Atzori, M. (2016). *Blockchain Technology and Decentralized Governance: Is the State Still Necessary?* Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2709713

Blanchette, J.-F. (2006). The Digital Signature Dilemma: To Preserve or Not to Preserve. *Annales des Télécommunications*, 61(7-8), 908-923.

Boucher, P. (2016, September 29). *What if blockchain technology revolutionised voting?* Retrieved from http://www.europarl.europa.eu/RegData/etudes/ATAG/2016/581918/EPRS_ATA%282016%29581918_EN.pdf

Bradbury, D. (2014). *coindesk*. Retrieved from How Block Chain Technology Could Usher in Digital Democracy: https://www.coindesk.com/block-chain-technology-digital-democracy/

Bralić, V., Kuleš, M., & Stančić, H. (2017). A model for long-term preservation of digital signature validity: TrustChain. In I. Atanassova, W. Zaghouani, B. Kragić, K. Aas, H. Stančić, & S. Seljan (Ed.), *INFuture2017: Integrating ICT in Society*, (pp. 89-113). Zagreb.

Chaum, D. (n.d.). Random-Sample Voting. Retrieved from https://rsvoting.org/whitepaper/white_paper.pdf

Cooper, D., Santesson, S., Farrell, S., Boeyen, S., Housley, R., & Polk, W. (2008). Retrieved from RFC 5280 - Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile: https://tools.ietf.org/html/rfc5280

Croatian parliment. (2002). *Narodne Novine.* Retrieved from Electronic signature law (NN 10/2002): https://narodne-novine.nn.hr/clanci/sluzbeni/2002_01_10_242.html

Croatian Parliment. (2002). *Narodne Novine.* Retrieved from Electronic signature Act (NN 10/2002): https://narodne-novine.nn.hr/clanci/sluzbeni/2002_01_10_242.html

(2015). *Data Dictionary for Preservation Metadata: PREMIS version 3.0.*

Doyle, M. D. (2002). *Patent No. United States Patent 6381696.*

Drescher, D. (2017). *Blockchain Basics: A Non-Technical Introduction in 25 Steps.* Frankfurt am Main: Apress.

Drummond, K. (2010). *Pentagon turns to brain implants to repair damaged minds*. Retrieved from Wired: https://www.wired.com/2010/05/pentagon-turns-to-brain-implants-to-repair-damaged-minds/

Dumortier, J., & Van den Eynde, S. (n.d.). Electronic Signatures and Trusted Archival Services. Retrieved 5 15, 2015, from http://www.expertisecentrumdavid.be/davidproject/teksten/DAVIDbijdragen/Tas.pdf

ETSI. (2016). Retrieved from ETSI EN 319 422: Electronic Signatures and Infrastructures (ESI); Time-stamping protocol and time-stamp token profiles: http://www.etsi.org/deliver/etsi_en/319400_319499/319422/01.01.01_60/en_319422v010101p.pdf

ETSI. (2016). ETSI EN 319 102-1: Electronic Signatures and Infrastructures (ESI); Procedures for Creation and Validation of AdES Digital Signatures; Part 1: Creation and Validation: http://www.etsi.org/deliver/etsi_en/319100_319199/31910201/01.01.00_30/en_31910201v010100v.pdf

European Comission. (2016, February 29). Questions & Answers on Trust Services under eIDAS. Retrieved December 29, 2017, from https://ec.europa.eu/digital-single-market/en/news/questions-answers-trust-services-under-eidas

European Parliament. (2014). *eIDAS*. Retrieved from https://www.eid.as/home/

Hanson, R. (2000). Shall We Vote on Values, But Bet on Beliefs? *Journal of Political Philosophy*, 1-40.

Herceg, B., Brzica, H., & Stančić, H. (2015). Digitally signed records - friend or foe? in: Anderson, K., Duranti, L., Jaworski, R., Stančić, H., Seljan, S., and Mateljan, V. (eds), *INFuture2015: e-Institutions – Openness, Accessibility and Preservation*, 147-150, Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb

Huminski, P. (2017). *The technology behind bitcoin could revolutionize these 8 industries in the next few years*. Retrieved 12 21, 2017, from Business Insider: http://www.businessinsider.com/8-applications-of-blockchain-2017-7

Ibrahimpašić, B., & Liđan, E. (2011). Digitalni potpis. *Osječki matematički list, Vol.10 No.2*, 139-148.

International Organization for Standardization. (2009). Retrieved from ISO/IEC 18014-3:2009 Information technology -- Security techniques -- Time-stamping services -- Part 3: Mechanisms producing linked tokens: https://www.iso.org/standard/50457.html

International Organization for Standardization. (2012). Retrieved from ISO 14721:2012 Space data and information transfer systems – Open archival information system (OAIS) – Reference model: https://www.iso.org/standard/57284.html

International Organization for Standardization. (2012). Retrieved 5 10, 2016, from ISO 16363:2012 Space data and information transfer systems – Audit and certification of trustworthy digital repositories: https://www.iso.org/standard/56510.html

International Organization for Standardization. (2012). Retrieved from ISO 16363:2012 Space data and information transfer systems -- Audit and certification of trustworthy digital repositories: https://www.iso.org/standard/56510.html

International Organization for Standardization. (2016). Retrieved from ISO 15489-1:2016 Information and documentation – Records management – Part 1: Concepts and principles: https://www.iso.org/standard/62542.html

InterPARES Trust Terminology Database. (n.d.). Retrieved December 28, 2017, from http://arstweb.clayton.edu/interlex/en/term.php?term=trustworthiness

Katulić, T. (2011). Razvoj pravne regulacije elektroničkog potpisa, elektroničkog certifikata i elektroničke isprave u hrvatskom i poredbenom pravu. *Zbornik Pravnog fakulteta u Zagrebu,* Vol. 61, No. 4, 1343-1344.

Lemieux, V. L., & Sporny, M. (2017). Preserving the Archival Bond in Distributed Ledgers: A Data Model and Syntax. *Proceedings of the 26th International Conference on World Wide Web Companion*, (pp. 1437-1443).

Merkle, R. C. (1980). Protocols for public key cryptosystems. *IEEE Symposium on Security and Privacy*, 122, pp. 122-134.

Mihaljević, M., Mihaljević, M., & Stančić, H. (2015). *Archival science dictionary. English-Croatian, Croatian-English.* Zagreb: FF Press.

Miroslav Krleža Institute of Lexicography. (2017). *Hrvatska encikolpedija*. Retrieved from digitalizacija: http://www.enciklopedija.hr/natuknica.aspx?id=68025

Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. Retrieved 11 21, 2015, from https://bitcoin.org/bitcoin.pdf

National Institute of Standards and Technology. (1999). Retrieved from Data Encryption Standard (DES): https://csrc.nist.gov/csrc/media/publications/fips/46/3/archive/1999-10-25/documents/fips46-3.pdf

National Institute of Standards and Technology. (2012). Retrieved from Digital Signature Standard (DSS): http://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.186-4.pdf

Nayuki. (2016, 1 4). Forcing a file's CRC to any value. Retrieved 10 5, 2017, from https://www.nayuki.io/page/forcing-a-files-crc-to-any-value

Rockwell, M. (n.d.). Retrieved from BitCongress - Process For Blockchain Voting & Law: http://bitcongress.org/BitCongress_Whitepaper.pdf

Santesson, S., Myers, M., Ankney, R., Malpani, A., Galperin, S., & Adams, C. (2013). Retrieved from RFC 6960 - X.509 Internet Public Key Infrastructure Online Certificate Status Protocol - OCSP: https://tools.ietf.org/html/rfc6960

Stančić, H. (2000). Digitization of documents. *2. i 3. seminar Arhivi, knjižnice, muzeji - Mogućnosti suradnje u okruženju globalne informacijske infrastrukture* (pp. 64-70). Zagreb: Hrvatsko knjižničarsko društvo.

Stančić, H. (2001). Archiving of Digital Documents. *4. seminar Arhivi, knjižnice i muzeji. Mogućnosti suradnje u okruženju globalne informacijske infrastrukture* (pp. 209-214). Zagreb: Hrvatsko knjižničarsko društvo.

Swan, M. (2015). *Blockchain: Blueprint for a new economy.* Boston: O'Reilly Media.

Underwood, S. (2016, November). Blockchain Beyond Bitcoin. *Communications of ACM, 59*. New York, USA: The Association for Computing Machinery.

What is an electronic seal? (n.d.). Retrieved December 28, 2017, from https://www.cryptomathic.com/products/authentication-signing/digital-signatures-faqs/what-is-an-electronic-seal

White House – National Science and Technology Council. (2016). Preparing for the future of artificial intelligence. Washington D.C. Retrieved 12 31, 2017, from https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

12. Appendix 1 – TRUSTER Preservation Model (EU31) – Case Study Questionnaire

This questionnaire was developed and used in order to make the case study results comparable. It was used in the case studies 1 and 2. It is divided in 11 sections with the corresponding questions.

QUESTIONNAIRE

The records in case are: _____

Interviewee: _____

Date of the interview: _____

Important note
Answers to the following questions should be
treated as classified (if applicable, please indicate): _____

1. Number of documents with expired certificates
Q: What is the number of documents with expired certificates being stored? Insight needed in order to see how much data we are dealing with.
A:

2. Number of requests for these documents with expired certificates
Q: How many requests were made for these documents? This can help us decide which documents are more important.
A:

Q: Are you dealing with documents that are accessible to users?
A:

3. Document importance
Q: Similarly to the previous question, are some of the documents more important, should any of the documents have a higher priority for preservation/migration etc.?
A:

4. Document age
Q: How much time has passed since document certificates expired, and what effect does that have on their potential recertification?
A:

5. Document timespan (how much time has passed since the certificates have expired)
Q: How much time has passed since certificates for certain documents expired? We consider this relevant because a longer timespan means more time for unauthorized access.
A:

6. In what way were the documents stored, what formats were used?
Q: Could you provide the insight into the way the documents were stored as well as on the formats in which they were stored? What technology was used and how outdated is the technology today? This is important in order to figure out a potential solution for the expired certificates.
A:

7. Document acquisition process
Q: In what way were the documents acquisitioned and what was the ingest process? Insight into this will give us more information on how the documents were kept and the possibilities for renewing their certificates.
A:

8. Document management of the files with expired digital signatures
Q: Who has access to the documents? Who manages the documents? How does this affect the trustworthiness of these documents? How does this affect potential recertification?
A:

9. Legal status of the expired signatures
Q: What criteria do the expired signatures have to meet in order to be considered as legally valid? Who decides on this?
A:

10. Business use
Q: Are the problematic records actually used for business purposes? If not, why are they kept, which parties are interested?
A:

Q: Are these records relevant to current or foreseeable high-value disputes and court proceedings?
A:

11. Long-term preservation
Q: Did you investigate the following as a solution for current or future long-term preservation problems? If yes, are they feasible?
A:

a. Re-validating of historical signatures using special software/hardware and/or third party services

   No / Yes → is it feasible Yes / No

b. Re-signing the records before the expiration of the signatures

   No / Yes → is it feasible Yes / No

c. Using e-notary services

   No / Yes → is it feasible Yes / No

d. Using trusted third party time-stamping and digital archiving services

   No / Yes → is it feasible Yes / No

e. Blockchain

   No / Yes → is it feasible Yes / No

f. Disposal of problematic records (changing legal requirements if necessary)

   No / Yes → is it feasible Yes / No

g. Validating records at the point of capture into trusted archival system, and afterwards trusting the system to ensure their integrity, usability and authenticity in time.

No / Yes → is it feasible Yes / No

h. Creating management system for records ensuring solid circumstantial evidence of their integrity and authenticity

No / Yes → is it feasible Yes / No

i. Other (please specify)

Q: Did you evaluate the cost of preservation (in various scenarios) vs. risks (such as fines for non-compliance, damages paid as a result of court cases etc.)?
A:

Q: How sure are you that the problem, once solved, won't repeat in time?
A:

a. Unsure
b. Somewhat unsure
c. Somewhat sure
d. Sure
e. Don't know