Publishing and using record-keeping structural information in a blockchain

Thomas Soedring[1], Petter Reinholdtsen[2], and Svein Ølnes[3]

[1] Oslo Metropolitan University, Oslo, Norway
tsodring@oslomet.no
[2] Norwegian Unix User Group, Oslo, Norway
pere@hungry.com
[3] Vestlandsforsking, Sogndal, Norway
svein.olnes@vestforsk.no

Abstract.
Purpose – To examine the role blockchain can play for record-keeping by exploring what information from a record-keeping system it is possible to publish to a blockchain. A credible approach is presented, followed by a discussion on both benefits and limitations.
Design/methodology/approach – The approach is a combination of theorised possibilities verified with practical software implementation. The basis for the work is relevant record-keeping and blockchain literature.
Research limitations/implications – The approach is beneficial where there is a record-keeping standard that has a clearly defined metadata model, and that also makes use of globally unique identifiers. Privacy legislation, e.g., GDPR, may limit the scope of an implementation of the approach.
Originality/value – The originality lies in presenting an approach whereby a record-keeping standard is analysed, separating structural and content information in order to publish structural information to a blockchain.
Keywords Record-keeping, Blockchain, Authenticity
Paper type Research paper

LaTeX-edition of article accepted for publication 2020-02-25 in Records Management Journal (DOI 10.1108/RMJ-09-2019-0056).

1 Introduction

Government agencies are subject to freedom of information (FoI) legislation that can ensure a minimum level of transparency and oversight. Government information is typically managed as records with associated metadata within a record-keeping system, and such systems are traditionally built on top of relational databases. It is possible to analyse the database of a record-keeping system and observe that there are various classes of metadata in play.

In particular, it is possible to see structural, content and process metadata. The structural metadata makes up the interconnected relationship of records (implemented, for example, with database foreign-keys), while the content metadata describes people, objects, and other information. Process metadata detail the underlying business processes and is typically implemented using status values.

There is a challenge here, in that all this metadata is intertwined, a necessary combination that ensures the efficient and effective operation of the record-keeping system. Taking a step back and unraveling the various classes of information may open for new ways to consider record-keeping and allow us to explore the integration of record-keeping and blockchain technologies.

Innovation commonly takes an idea and converts it into something that creates value for somebody. The advent of blockchain technology is an innovation that opens up new possibilities to rethink the concept of trust, however as blockchain technologies are still in their infancy, it may still be unclear where the actual potential value lies. As blockchains have a high degree of immutability, they are an exciting approach to preserving originality and proof that information remains unaltered.

The underlying premise for this work is a cross-disciplinary approach to the issue of trust in government records by exploring the integration of record-keeping and blockchain technologies, where structural metadata from a record-keeping database is published to a blockchain. Our research questions are: How can structural information be extracted from a record-keeping system and stored on a blockchain, and what benefits does this give?

The remaining article is structured as follows. First, there is a review of related research. After that, the methodology and approach for the work are discussed. The blockchain concept is considered before the Noark record-keeping standard is presented, with particular attention to the properties that apply to blockchain integration. The approach is then presented, and details of the integration of blockchain and record-keeping are described. The potential benefits and limitations of the approach are discussed, followed by a conclusion.

2 Related research

The solution proposed here is comparable to the solution described in (Lemieux and Sporny, 2017). There are apparent similarities between both solutions, requiring a unique identification of records, and a bond between records and documents. They extend their approach to a description with linked data, while the approach we propose is broader and based on an existing record-keeping standard. Further, our approach aims at a much tighter integration between blockchain and the record-keeping system. Their approach is more general, discussing issues regarding the archival bond and blockchain.

A blockchain is not limited to the concept of ledgers, even though they are often associated with ledgers. The work presented here builds upon the under-

lying concept of blockchain, rather than distributed ledgers, which is in contrast to Lemieux (2017*a*), where the focus is more related to the ledger concept.

Lemieux (2017*b*) also discusses potential preservation issues for record-keeping systems built on top of blockchain as well as initiatives occurring from the archival perspective. Preservation is discussed in this article, but our discussion is simplified when compared to Lemieux (2017*b*).

Lemieux et al. (2019) have also undertaken a significant analysis of various issues relating to the role of blockchain in both record-keeping and preservation. Findlay (2017) lays forth the concept of trust in record-keeping and a call to action to integrate the two. Our approach sheds particular light on implementation issues when integrating blockchain and record-keeping.

It is also feasible to solve the authenticity issue using trusted time-stamping (Adams et al., 2001). Trusted time-stamping is built upon public key infrastructure, allowing for the signing of records by a trusted third party. Verification is possible using the public part of a public/private key pair of the trusted third party. A trusted time-stamping approach has some benefits when compared to blockchain. Many services (e.g.*https://originstamp.org*, *https://freetsa.org*) already provide trusted time-stamping, and there should be a relatively low investment cost associated with the technology.

3 Methodology and approach

ISO-15489-1 (2016) defines records management as the "field of management responsible for the efficient and systematic control of the creation, receipt, maintenance, use, and disposition of records, including the processes for capturing and maintaining evidence of and information about business activities and transactions in the form of records".

The work here pays particular attention to two parts of the record management definition. The first concerns itself with the concept of value as evidence and lays forth an approach that may potentially increase the value as evidence of records. The second concerns itself with information about business activities and transactions and exposes such information about activities along with integrity information to a blockchain.

The Norwegian record-keeping and preservation standard, Noark 5 (Arkivverket, 2018) is used as a basis for experimentation. The Noark 5 standard is based on both (ISO-15489-1, 2016) and (ISO-16175-2, 2011). Lemieux (2016) notes that the specific design of solutions that integrate blockchain and record-keeping requires further research and experimentation. The work here picks up on this issue and demonstrates a unique approach to integrating the structure of a record-keeping database with a blockchain.

The presented approach is verified using two free and open-source software implementations. The first is an implementation of the Noark standard (Soedring Thomas, 2013), while the other is a simple blockchain implementation (Thomas, 2013), that can also export the blockchain to a preservation friendly format. The Noark record-keeping system has an event handler that captures all

Create, Read, Update, and Delete (CRUD) events, and information about these events is subsequently published to a blockchain.

4 Blockchain technology

Modern blockchain technology traces its roots to a white paper detailing the Bitcoin electronic cash system (Nakamoto, 2008). Bitcoin has an extraordinary history, and its influence in the financial sector is likely to be felt far into the future. The blockchain concept also has the potential to be a disruptive technology in other sectors beyond the financial sector.

A blockchain is a relatively simple construct, as shown in 1. The first block starts with a time-stamp and some data. A checksum covering the contents of the block is calculated and stored before the block is finalised. The second block consists of the checksum from the first block, along with the current timestamp and the current blocks' contents. A new checksum is calculated and stored before the block is finalised. This approach continues, and new blocks are added to the ever-growing chain, resulting in a tamper-evident chain of blocks, or a blockchain. A new block will always contain the checksum of the previous block, thereby creating a bond between the blocks. If a block has been manipulated, an analysis of the chain will quickly identify the problematic block.

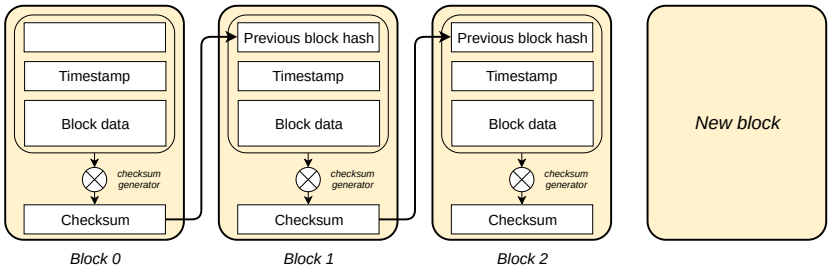

Fig. 1: A simple blockchain showing how the checksum for a block covers both the data in the current block as well as the checksum from the previous block.

Depending on the implementation scenario, the chain of blocks is expanded to include ledgers, in order to record transactions. A blockchain is normally distributed, resulting in increased authenticity as there are multiple copies of the chain verifying the overall blockchain. However, this increases complexity as it must be possible to reach a consensus with regards to the validity of the chain and how to add new blocks.

One of the promising use-cases of blockchain is to support "smart contracts". A smart contract is software that automatically executes the terms of a contract upon fulfilment of a predefined condition. Once all parties agree that the conditions of a contract are met, payment can automatically be made. Such an

approach increases transactional efficiency, eliminating the need for a middleman while maintaining a high level of transparency (Crosby et al., 2016).

The introduction of new technology is often associated with a certain amount of hype, and blockchain is no exception to this. There is a requirement for a sobered approach when considering the implementation of a technology like blockchain, as for government, any long term implementation will come with significant financial commitments. A government should avoid a rush to blockchain as a generic solution to trust issues, rather they should invest time to understand what problems they are experiencing that blockchain may solve and whether such problems can be solved using other technologies.

Gartner notes that blockchain is now in the "trough of disillusionment" within the Gartner hype cycle (Litan and Leow, 2019), which is a sign that hype-interest in the technology is waning as some experiments and implementations fail to deliver. It can also be a sign of maturity, as the technology carves itself a niche market.

5 Record-keeping based on Noark

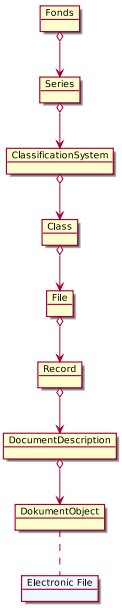

Fig. 2: Commonly used Noark record-keeping structure

Norway mandates by law (Kulturdepartementet, 2018), that government agencies undertake record-keeping based on a national record-keeping standard

that defines a life cycle approach to the management of records, from the time of their inception to an eventual disposition.

The standard was first used in the early 1980s and through various versions aims to follow changes in laws and regulations, information technology, and record-keeping principles and practices. The Noark standard is currently at version 5.5 and consists of both conceptual as well as technical descriptions.

A Noark record-keeping system is a documentation archive of the public administration and functions as an archive of executive authority, as well as a correspondence archive for public administration. Storing government records based on a standardised approach also makes it easier to expose records to the public, thus enabling easier access for citizens when undertaking FoI requests (Hagen Sataslåtten, 2017).

The Noark standard defines a metadata model that manages central metadata relevant both to record-keeping and preservation. The commonly used interpretation is shown in Figure 2. At the highest level, the *Fonds* and *Series* entities typically define an organisational context for records, while the *ClassificationSystem* and *Class* entities typically define a functional context. The *File* and *Record* entities define a transactional context that includes a document context, defined by the DocumentDescription and DocumentObject entities.

Noark is also *preservation ready*, in that it has standardised descriptions defined as XSD (XML Schema Definition) schemas defining the structure of data-extractions for preservation purposes. Another feature of the Noark 5 standard is that each instance of a Noark 5 object (a row or tuple in database terminology) must be unique within the organisation, not just within the record-keeping system. Each Noark object is identified by a field called systemId, and there is a requirement to generate systemId values according to a standard called UUID or universally unique identifiers (Leach et al., 2005). A UUID approach should result in identifiers with an extremely low probability of duplicates, even globally.

6 Combining Noark and blockchain technology

It is possible to consider the underlying database of a record-keeping system from multiple perspectives. From a user perspective, a user may only be interested in seeing records presented in the correct visual context, as users typically do not concern themselves with the details of how the records are stored. A software developer may be concerned with the underlying data model and how records are processed, while a database administrator may be concerned with performance issues and may undertake optimisations. A data quality analyst may try to understand the quality of data in various contexts, e.g. examining individual records or relationships between records.

Another consideration when examining records in the database is to distinguish between structural and content information. In the case of Noark, foreign keys typically link objects together in parent-child relationships. To a certain degree, it is possible to say that this linking establishes an archival bond (Duranti, 1997) for the records.

The Noark record-keeping metadata model is a relatively straight forward model that consists of a set of defined objects with associated metadata. As each object is associated with at least one parent object, it is possible to traverse the record-keeping structure from any given point. Some objects have references to siblings or other related objects, making the entire structure a graph; however, for the sake of brevity, the structure is shown here as a tree. Listing 1.1 describes the structure depicted in2 in XML. The listing shows the embedding of the various objects that make up the record-keeping structure, resulting in a baseline hierarchical structure. It also shows the assignment of UUID values to the systemId field, as well as a minimum amount of provenance information linking the records to an organisation. For simplicity, Oslo Municipality is used here as an example.

XML is used to describe metadata and data as it allows for a consistent and straightforward presentation of the Noark record-keeping metadata model.

```
<?xml version="1.0" encoding="UTF-8"?>
<fonds>
 <systemId>cd4861b4-a861-4d9f-bda0-8c0683d6e60a</systemId>
 <title>Oslo Municipality</title>
 <series>
  <systemId>395a57bd-db07-461a-8f14-57db7f43c733</systemId>
  <classificationSystem>
   <systemId>d2450bb7-498c-497d-af10-3f198887dd17</systemId>
   <class>
    <systemId>1b05b631-76c4-4893-b9eb-adf03a8f45a2</systemId>
    <file>
     <systemId>d3134332-30b1-42ec-924e-ecfe80c2bac3</systemId>
     <record>
      <systemId>167ccf56-082e-4d7a-93cf-f9ab619b3537</systemId>
      <documentDescription>
       <systemId>a80ac37c-4f9a-4202-a8bf-7885609774be</systemId>
       <documentObject>
        <systemId>cf00c46a-51f2-f210-f1c1-b2851931442b</systemId>
       </documentObject>
      </documentDescription>
     </record>
    </file>
   </class>
  </classificationSystem>
 </series>
</fonds>
```

Listing 1.1: Basic Noark record-keeping structure described with XML (translated to English, Noark use Norwegian tag names)

One approach to solving the problem of integrating record-keeping and blockchain consists of taking the structural information shown in Listing 1.1 and publishing it to a blockchain. Such an approach will ensure that the structural information that makes up the record-keeping database is mirrored to a blockchain. As the Noark record-keeping structure establishes an archival bond, the solution then publishes the structure of the archival bond to a blockchain, without publishing data.

On its own, publishing structural information serves a limited purpose. From a transparency perspective, publishing a copy of the record-keeping structure on a blockchain merely provides a public notice that record-keeping is taking place. If a particular government agency is notorious for not undertaking record-keeping, then the publication of such structural information can raise questions detailing the running of the organisation.

Within the transactional context of the record-keeping process, it may be possible to increase the level of transparency by publishing relevant status values and time-stamp information in addition to the structural information. Listing 1.2 details the creation and finalised times of a particular case file. Furthermore, a record associated with the case file shows the existence of a PDF-file. This record states that the document has an "Approved" status. For the sake of brevity, a lot of the other metadata that would be present is left out.

```
<?xml version="1.0" encoding="UTF-8"?>
<file>
 <systemId>d3134332-30b1-42ec-924e-ecfe80c2bac3</systemId>
 <createdDate>2019-08-12T12:14:03+0200</createdDate>
 <finalizedDate>2019-09-23T14:19:12+0200</finalizedDate>
 <caseStatus>Finalized</caseStatus>
 <record>
  <systemId>167ccf56-082e-4d7a-93cf-f9ab619b3537</systemId>
  <createdDate>2019-09-191T15:59:21+0200</createdDate>
  <archivedDate>2019-09-20T08:25:42+0200</archivedDate>
  <recordStatus>Approved</recordStatus>
  <recordType>Outgoing document</recordType>
  <documentDescription>
   <systemId>a80ac37c-4f9a-4202-a8bf-7885609774be</systemId>
   <documentStatus>Document is finalized</documentStatus>
   <documentType>Letter</documentType>
   <documentObject>
    <systemId>cf00c46a-51f2-f210-f1c1-b2851931442b</systemId>
    <format>PDF</format>
    <checksum>ca007a06595dded947805ecbe4d5374c7a15166e</checksum>
   </documentObject>
  </documentDescription>
 </record>
```

```
</file>
```

Listing 1.2: Showing the combination of structural and process information described with XML

The goal behind this approach is to publish information about the record-keeping process along with structural information, without publishing content information or any personal information. It is possible to extend this approach with further metadata, that may also be acceptable to publish publicly. Listing 1.3 shows additional content, where it has become clear that the shown records are about a building application that has been handled by the local municipality.

```
<?xml version="1.0" encoding="UTF-8"?>
<file>
 <systemId>d3134332-30b1-42ec-924e-ecfe80c2bac3</systemId>
 <createdDate>2019-08-12T12:14:03+0200</createdDate>
 <finalizedDate>2019-09-23T14:19:12+0200</finalizedDate>
 <caseStatus>Finalized</caseStatus>
 <caseNumber>2019/00146</caseNumber>
 <title>Application to build conservatory Drammensveien 1 Oslo</title>
 <record>
  <systemId>167ccf56-082e-4d7a-93cf-f9ab619b3537</systemId>
  <createdDate>2019-09-191T15:59:21+0200</createdDate>
  <archivedDate>2019-09-20T08:25:42+0200</archivedDate>
  <recordStatus>Approved</recordStatus>
  <recordType>Outgoing document</recordType>
  <documentDescription>
   <systemId>a80ac37c-4f9a-4202-a8bf-7885609774be</systemId>
   <documentStatus>Document is finalized</documentStatus>
   <documentType>Letter</documentType>
   <documentObject>
    <systemId>cf00c46a-51f2-f210-f1c1-b2851931442b</systemId>
    <format>PDF</format>
    <checksum>ca007a06595dded947805ecbe4d5374c7a15166e</checksum>
   </documentObject>
  </documentDescription>
 </record>
</file>
```

Listing 1.3: Showing further content information described with XML

Listings 1.1, 1.2, and 1.3 show an evolution in the type of data that it is possible to store on a blockchain. Listing 1.1 shows mainly structural information along with some provenance information describing the government organisation. The organisation's record-keeping structure is visible in the blockchain, and the blockchain only shows the creation rate of records. Listing 1.2 adds additional content information where the record-keeping process becomes visible through

various status values, while Listing 1.3 publishes even more content information describing the context for a given case file.

The amount of information that it is possible to publish will depend on the government agency. It is likely that it is possible to publish most data from records for building applications, as there is an inherent requirement to keep such information public. Conversely, limited or perhaps even no information is publishable from case files handled by a child protection services agency. Depending on the business area, there will likely be ranges of data that is possible to publish. There are also government areas that have limited requirements to follow record-keeping regulations. Intelligence agencies, for example, may wish to be kept outside of the scope of such a solution. A potential reason for this is that foreign adversaries may observe record-keeping in the blockchain to see what kind of global and local events cause the registration of data. A potential solution, in such a case, is to publish only a subset of the data shown in Listing 1.1, e.g., excluding the time-stamps, and to publish at monthly intervals.

An important consideration that is a requirement to make this approach feasible is that it must be possible to trace a data object (that has a systemId) to its parent object. Linking to a parent object is possible by including the parent systemId in the data section of a blockchain block, or by adding it to a block header. Such an approach builds an application layer over a blockchain in the same way that bitcoin builds an application layer over a blockchain.

The approach presented here makes use of a blockchain, without laying forth a particular requirement for any given consensus model, like the "proof-of-work" model used by Bitcoin. The final choice of a consensus model is left to be decided based on the requirements of an actual implementation for a given government. As the given context refers to government records, a "proof-of-authority" (Gavin, 2015) may be an applicable approach when deciding a consensus model.

7 Applicable use-case

A relevant use-case that can shed light on how the proposed approach could work from a practical perspective is how the police secure evidence, in particular how the police establish a chain of custody when collecting evidence for later use in a criminal court case. The O.J. Simpson trial exemplifies this issue as Simpson's defence was able to sow doubt about the police chain of custody covering a pair of gloves found at the crime scene. During the trial, Simpson struggled to get his hand in the glove, quipping that they were too tight (Evans, 2003, p. 227).

A modern approach to handling evidence could be to first collect a unit of data and metadata of electronic evidence at the crime scene. A checksum covering this data is stored on a blockchain with an identifier, that can be used to verify the integrity of the information at a later time.

- Checksum of image content (4ee263eb7e5fbaed824fb3df0a8db2e85b4eea9c)
- Camera Information (Nikon D5100)
- GPS Coordinates (59.91111, 10.75278)

– Time (2019-09-02T16:17:15+0200)

```
<?xml version="1.0" encoding="UTF-8"?>
<evidencefile>
 <systemId>39cfeb90-0958-4af9-aa43-f7619162f079</systemId>
 <createdDate>2019-09-02T16:18:17+0200</createdDate>
 <record>
  <systemId>0b587313-dbf3-4256-8977-ea9f703bedd2</systemId>
  <createdDate>2019-09-191T15:59:21+0200</createdDate>
  <location>59.91111, 10.75278</location>
  <documentDescription>
   <systemId>9ec05cb3-4cad-48b8-8003-e147cd0edc75</systemId>
   <documentObject>
    <systemId>1e38d84b-80fe-4606-91a7-551ac915b728</systemId>
    <format>jpeg</format>
    <checksum>4ee263eb7e5fbaed824fb3df0a8db2e85b4eea9c</checksum>
   </documentObject>
  </documentDescription>
 </record>
</evidencefile>
```

Listing 1.4: Showing structure and content of evidence metadata marked up with XML

Listing 1.4 shows the structural and content information of an evidence file that contains information about a single image. All metadata related to the collection of evidence of a crime scene is collated within the same file. However, the location field pinpointing the exact location of a photo may be problematic from a privacy point of view. Consider a high profile murder case, that has enraged a high proportion of the citizenry. If a blockchain shows that the police collected evidence at a particular location, the citizens may suspect that the person at the location is the perpetrator and react negatively.

A possible solution to the problem is to store the geolocation using the EXIF specification (JEITA, 2019) as part of the image, and not as metadata directly searchable in the record-keeping system. It is then possible to use the image checksum to validate both the image and the geolocation while keeping the location out of the blockchain. Storing data on a blockchain may quickly lead to a scenario where too much information is published, and care must be taken to avoid this.

The combination of a blockchain and Noark is shown in Figure 3. Here it is possible to see the first and second block in the blockchain. In the first block, a time-stamp, along with the relevant Noark metadata required to show the creation of a File object, is shown. A checksum covering the time-stamp and the Noark File data is calculated. In the second block, the checksum from the previous block, along with the time-stamp for the current block and the relevant

Noark metadata corresponding to the creation of a Noark record object, is shown. A checksum covering these three is then calculated and can be used as the ingress checksum for the next block. Note the parent element in the second block points back to the Noark element in the previous block. Applying this approach requires that the systemId field in the record-keeping system must also be immutable.

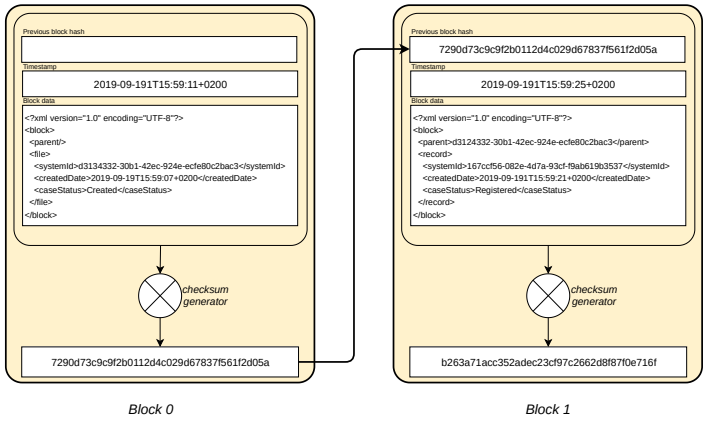

Fig. 3: Two blockchain blocks containing Noark information

8 Discussion

The blockchain integration approach makes use of Noark's standardised metadata model, where nearly all objects have at least one parent; therefore, it is possible to trace the interconnected nature of records through the blockchain. The record-keeping structure is not published directly to the blockchain, rather portions of the record-keeping structure are published in response to events that occur in the record-keeping system. As such, the blockchain copy functions more as a change-log detailing the history of records, but it should be possible to build a local copy of the record-keeping structure from the version in the blockchain. This answers our first research question of how to extract the structural information of a record-keeping system and publish it to a blockchain.

Our second research question asked what benefits this approach can have. Publishing such information to a blockchain can give records a new innovative dimension of usability, while simultaneously increasing both transparency and integrity. The approach has both benefits and limitations, ranging from additional levels of published information to base FoI requests upon, to analytical insights of the record-keeping process, to the blockchain acting as a master record of sorts. The limitations are that it is unclear how much information it is possible to publish without infringing on someone's right to privacy.

8.1 Transparency

Providing the public with an ability to traverse the record-keeping structure through a blockchain will significantly improve the level of transparency for government records. The current Norwegian approach to transparency for government record-keeping, where information about documents is published to a public correspondence journal, exposes limited information about the process that created the document. The approach presented here should result in increased transparency as an insight into the process the document was subjected to can now be shared with the public.

There have been a few cases in the Norwegian media over recent years that provide support that the presented approach may be necessary. One instance (Bakke, 2017), relates to how the record-keeping system belonging to the Norwegian Police Directorate inadvertently prevented the publication of information about documents. Documents that detailed a high level of overspending for a particular project were hidden from the public view because status values that would have resulted in the publication of the documents' existence were incorrectly set. Publishing metadata about processes and documents to a blockchain would afford the public an ability to trace through the public version of the record-keeping structure belonging to the police directorate, exposing any instances where documents were failing to follow proper routines.

Another instance (Christina, 2018), is where an employee working for the local planning office of Drammen municipality abused roles she had obtained in the record-keeping system. She had acquired a management level role allowing her to log in to the record-keeping system as both a case handler and as a manager, and was able to approve building applications, without proper oversight. A public version of the record-keeping structure on a blockchain that also identified users with a unique value could quickly expose situations where a case handler self-approves a case file.

It is possible to establish a high level of granularity in the relationship between a blockchain and the record-keeping system by publishing process information governing the process documents have been subjected to. Capturing CRUD events in addition to status values can provide a comprehensive description of a document's life-cycle, without exposing any privacy infringing information. It is, however, unclear if it is possible to publish the identifiers of government agents acting in an official capacity to a blockchain. A government agent could attempt to request the deletion of such published information with a GDPR (Commission, 2016) 'right to be forgotten' request. The immutable nature of a blockchain is incompatible with such a request.

8.2 Transparency analytics

The blockchain copy of the record-keeping structure can be used as a basis for analytics, allowing for comparative analysis within and across record-keeping systems. For example, two similar-sized municipalities should, in theory, have a

comparable record-keeping structure with a similar quantity of records. Publishing the record-keeping structure of both municipalities on top of a blockchain allows for independent analysis of the record-keeping process. If one municipality has a significantly smaller record-keeping structure, it is possible to question why this is the case.

Publishing the record-keeping structure to a blockchain can also, to a degree, be used by government auditing agencies to ensure that record-keeping is taking place per regulation. One of the responsibilities of the National Archives of Norway is to audit the record-keeping process in government agencies and ensure it occurs according to laws and regulations.

Using the police directorate example from the previous section, those with responsibility for the record-keeping process may have been in a better position to undertake an analysis of the record-keeping process if there was readily available access to structural and process information. However, from their perspective, there may be no requirement to publish such information publicly; instead, it could be part of the internal daily routines.

The examples here point to the fact that it is possible to publish more information from a record-keeping system, than that which is being published today. Incidentally, publishing such information on a blockchain highlights this possibility. The increased possibilities for citizen or journalist analytics are a side effect of increasing transparency. It is unlikely, though, that the requirement to publish structural and process information would occur on its own. The innovation potential in blockchain affords new possibilities and ways to utilise the record-keeping systems to increase transparency.

8.3 Preservation authentication

The approach presented here can also have an implication when interpreting the concept of a master record. A master record can be seen as the definitive copy of a record held by a government agency, and to a degree, is the record that maintains value as evidence. University grade transcripts can illustrate the concept of a master record, as the official university copy is typically the one that holds value as evidence. MIT has, for example, created a blockchain-based solution for validating university transcripts (Jirgensons and Kapenieks, 2018), a fact that illustrates that blockchain technology already can play a vital trust role in records.

The linear nature of time is evident in a blockchain, and this is an essential additional integrity mechanism. In a poorly secured record-keeping system, it may be possible to back-date records by directly editing records in the database and covering up the attempt by manipulating log files. Employing a third party trust mechanism, like a blockchain, will ensure that it is possible to identify any manipulation attempts in the record-keeping system.

The concept of trust, when compared to today's approach, can be augmented by distributing the mechanism to determine the integrity of records between a blockchain and any system that has a copy of the records. This is achieved by storing metadata about the document along with any integrity data, e.g., a

checksum, in the blockchain, while the actual document or record is stored in the organisation's record-keeping system.

It may also be possible to preserve integrity when migrating data from one system to another. It is possible to maintain the integrity of migrated data between systems via the blockchain copy, as integrity is no longer constrained to the record-keeping system. This may also play a role when migrating the record-keeping system for long term preservation. The Noark standard defines a set of XSD schemas to validate an extraction of the records from a Noark record-keeping system. The blockchain copy can play an essential role in maintaining integrity during a migration process for preservation purposes. When an archive institution receives a copy of the extracted records, this can be checked against a blockchain version. Here it will be possible to observe if all records have been extracted or not and to verify that documents are authentic.

Validating an extraction of records from the record-keeping system against the copy that the archive receives is an important issue. (Høiaas et al., 2016) undertook an analysis of an extraction from a Noark 4 record-keeping system to see if there were any notable problems. The analysis showed that 49% of the electronic documents within the original record-keeping system were missing. It is believed this was a result of a software error. The work highlights the importance of validating records and stands as a reminder that migration has inherent risks.

8.4 Preserving the chain

If blockchain technology plays a vital role in establishing trust for record-keeping, then it will one day become an issue for preservation. At its simplest, a blockchain can be seen as a list of interrelated blocks of information, and it is a relatively straight forward process to migrate blockchain blocks to a preservation friendly format like XML or JSON. However, there may be some IT-ecosystem dependencies that may have a significant impact on the future technical cost of preserving a blockchain. As such, it is essential to keep any government blockchain infrastructure to validate records or documents, as simple and straightforward as possible.

Considering the Gartner hype-cycle, potential preservation problems when blockchain technology reaches the plateau of productivity should become evident. Early attempts that are subsequently abandoned may require an investment to preserve the blockchain infrastructure in some form or to declare that it is not possible to validate records from early implementations.

One can expect that government-backed blockchain for areas like smart contracts will have a very long life span, likely to be in terms of decades. If blockchain technology becomes essential to the management of trust, it becomes essential to the running of government, and as such, may never become a preservation issue. However, there is also a temporal issue at play here. In 2089, will we care about the authenticity of records from 2019? It is likely that, at some point, records will lose the requirement for trust as a society may hold little value in the ability to prove the authenticity of older records.

The migration of records or documents will naturally challenge any checksum or signature for objects held in an archive. Any known change to a bitstream of a preserved object must result in a new entry in the blockchain for that object, preferably with a link back to the block of the original object. Before a migration, the original record or document is validated with its corresponding blockchain information. The result of this migration is also recorded along with the migrated object.

It is also possible to preserve a blockchain for its historical value even if it has no real use anymore. The actual process of preserving a version of a blockchain for public use is straight forward, as there is only a requirement to read information from the blockchain, not write to it. The chain can easily be stored in a relational database with a simple web-facing API. For internal reference use, more optimised approaches can be considered. Such an approach requires that the institution must be able to independently prove the authenticity of the blockchain.

8.5 Handling privacy

Within the presented approach, there are potential privacy issues as variations in the type of data published is increased. There is an inherent risk when sending information from a record-keeping system to a blockchain that personal information will leak to the chain. The immutability of a blockchain makes this is an issue that must be addressed earnestly. Wrongly registered data in a record-keeping system is relatively easily corrected. It is not possible to correct data wrongfully published in a distributed blockchain, and the chain may need to be purged, ultimately questioning the integrity of the chain.

The premise for the approach presented here, to integrate record-keeping and blockchain, rests on the fact that only structural and process information is published to the chain as described in Listings 1.1 and 1.2. Listing 1.3 and 1.4 begin to point indirectly to personal information and quickly show that privacy issues arise.

This issue can become a potential problem with regards to privacy focused laws like the GDPR (Commission, 2016), as the blockchain may not comply with GDPR if personal information is exposed there. Care must also be taken to ensure that such structural information does not indirectly leak information through other systems or sources. A hypothetical example is a situation where the local media reports on a terror attack where a local citizen is seriously injured. The injured person may subsequently apply to the municipality for help, and it may be possible to logically conclude that the structural information published in the blockchain relates to the injured person residing in the municipality.

8.6 Rethinking trust

An electronic record is considered intact and uncorrupted if its "identity is clear and the message that it is meant to communicate in order to achieve its purpose is unaltered" (Duranti and Blanchette, 2004, p .216). Record corruption does

not need to be a result of malice. Instead, it may be a result of the idiosyncrasies of systems. An integration between blockchain and record-keeping may help detect electronic record corruption as the blockchain can provide independent verification of record contents. Our approach requires a clear identification of records, so information about records stored in the blockchain will have a clear identity. In our implementation, a UUID approach is used. The immutability of the blockchain can deal with the message of a record being unaltered. At the simplest level, adding checksums governing the record can help detect if the record remains unaltered. This will, however, depend on how much authenticity metadata is placed within a blockchain.

There may be little benefit in monetary value from storing integrity information on a blockchain; instead, society may see increased social trust. It may be difficult to see the need for such technology if there already is a high level of citizen trust in government. Northern European citizens have high a level of trust in various public institutions, and in particular, have high confidence in the police and the judiciary (Kleven, 2016), a fact that may decrease the pressure for the uptake of blockchain technology for public record-keeping. If a government organisation already has a high level of trust, then it is likely that the organisation will not see any additional value from integrating its record-keeping systems to a blockchain. Blockchain technology does not magically solve all trust issues, but rather enables new trust architectures (Werbach, 2018) for society, that can also be used to independently verify parts of the record-keeping process. In such a case, employing blockchain technology is more about utilising modern trust infrastructures in order to increase transparency.

When considering the Gartner hype cycle, many have likely searched for a question to which blockchain technology is the answer. Is the approach presented here also an answer looking for a question? The answer may lie in looking at what additional value blockchain technology provides for record-keeping, and what is possible beyond a trusted time-stamping approach. Trusted time-stamping ensures the integrity of individual objects, e.g. documents and records, while blockchain technology can ensure the integrity of an entire record-keeping system and related organisational processes. The distinction is vital as trust is apparent at two different levels. An approach based on blockchain technology will not provide a binary answer with regards to the authenticity of a record-keeping system. Instead, it affords us an additional level of integrity not seen before. Various content information, e.g. checksums stored in the blockchain, may give binary answers to the question of trust for documents, but so too can trusted time-stamping. As trusted time-stamping is often based on public key infrastructure, it will likely achieve a higher level of trust for individual records. There is no clear answer here, as the question ultimately is, what level of trust do we seek to achieve.

Another consideration is that the record-keeping systems and archives may become targets by foreign adversaries wishing to sow discord. It is known that Russia tried to exert influence on the result of the 2016 US presidential election, in order to sow dissatisfaction within the political system (Mueller, 2019, p .14).

As the digital society grows, more and more records will be born-digital in electronic record-keeping systems, and there will likely not be paper-based versions that, in a final question of doubt, can verify authenticity. Cryptographic signature mechanisms have replaced rubber stamps, but such signatures typically only cover electronic documents. Entire databases of records likely exist without sufficient integrity mechanisms. The question can be asked, are governments ready, willing, and able to handle coordinated attacks on their records?

Electronic record-keeping plays an essential yet invisible role in so many aspects of our daily lives. Electronic systems today are used to quickly verify that a particular driver's license is valid or that vehicle tax, registration, and insurance are up to date. A hostile foreign agent may attempt to sow discord within a society by tampering with citizens' rights. Some examples of attack vectors include:

- Driver license set to be invalid, causing problems during routine checks
- Incorrect changes to tax information resulting in incorrect back tax
- Deletion of information about work history that can delay pension rights.

Society may find itself more dependent on the ability to verify the authenticity of records, challenging the concept of societal trust. Blockchain technology may show itself not as a new luxury technology for record-keeping, but rather the tool by which we can trust the integrity, authenticity, and reliability of records

8.7 Making the case

Integrating blockchain and record-keeping affords society new ways to consider trust in government. Ultimately government organisations will need to see an inherent benefit in the approach; otherwise, they will not embrace it. The use of blockchain technology, however, could become politically mandated if politicians see the benefit of increasing or maintaining trust in government organisations, and are willing to bear a cost for maintaining a distributed blockchain. It is likely that the integration of blockchain technology and record-keeping will be based on an existing investment in blockchain technology for other areas of government responsibility, rather than investing in blockchain technology solely for record-keeping. It is possible to introduce the approach presented here as a requirement in the Noark standard and only limit the data to structural and process information. Rolling out such an approach can be done with minimal disruption to the government organisations.

A question can be raised, whether or not a blockchain is required when good record-keeping practices are followed. A similar issue exists, for example, when government contracts out road maintenance. It could be argued that with a good contract and a reliable service provider, the task of checking the quality of the road and its need for maintenance can be assigned to the service provider, and there is no need for the organisation buying the service to spend resources on verifying that the road is properly maintained. However, any sound system needs checks and balances, and it is essential to "trust, but verify" as the Russian proverb used by Ronald Reagen goes.

A similar case can be made in terms of record-keeping. How is it possible to determine that good record-keeping practices are being followed? The systems are often closed to access from the public. A record-keeping system in an average municipality in Norway can have many hundred users, and it can be challenging to measure how well practices are being followed.

In Norway, the Noark-based record-keeping systems publish a public journal on a daily basis, but the idiosyncrasies of a system show that it is possible that information that should be made public, actually is not made public (Bakke, 2017). There is definitely a case to be made to publish structural and process information to increase transparency and thus trust in government record-keeping. In some ways the approach presented here takes parts of the question of trust in a government organisation and lifts it up outside of the organisation and up to a higher level in society. It can be argued that in some ways trust in government organisations is based on blind faith, and is challenged when a scandal is blasted across the front pages of national media. The ability to verify is an important foundation to establish trust. Blockchain technology can be one such infrastructure to manage the documentation for trust. We can increase trust in government, because we can verify it.

9 Conclusion

A government has a particular responsibility in managing the authenticity of records for citizens. Land ownership registries are an example of this where it must be possible to trust that a government maintains correct and up to date information. Blockchain technology may afford government new capabilities to managing authenticity, beyond that which is possible to achieve today, through an ability to have independent third party validation of records.

The approach presented here discusses the concept of authenticity by examining an existing record-keeping standard and exploring how this record-keeping standard can see better integration with blockchain technology. The Norwegian record-keeping standard, Noark, is used in this regard as it has a standardised metadata model. An analysis of this model shows that it is possible to distinguish structural information from content information and publish structural information to a blockchain.

Both benefits and limitations of the approach are discussed. Publishing structural and process information is likely unproblematic, but as soon as content information from case files is published, there may be issues concerning privacy and legislation such as GDPR. No clear answer to this problem is provided; instead, the problem is identified and left for further discussion. GDPR and blockchain are discussed both in (Hofman et al., 2019) and (Lemieux et al., 2019).

The approach introduces new ways to handle freedom of information, and shows that a record-keeping structure published to a blockchain affords a new level of insight into the government record-keeping process. The approach enables new forms of citizen driven analytics providing insights into the entirety of the record-keeping process. Publishing such information may help detect sit-

uations where information is prevented from being published, and may even be a deterrent to corruption.

It is possible to deal with some of the problems discussed here without the need for blockchain. Individual records and documents can be verified using trusted time-stamping. Freedom of information infrastructure exists in Norway, where information about documents is published. It is feasible to deal with the issue of analytics through a better understanding of the underlying database or other analysis. Blockchain technology is not a requirement to improve these issues, but, interestingly, the combination of blockchain and record-keeping may result in new innovative solutions. This a classic example of how a disruptive technology solves a problem that many did not know existed.

The application of blockchain technology to record-keeping has the potential of increasing social trust by adding an independent layer of authenticity to record-keeping. If there is a high level of trust in public agencies, there may not be much of a push for blockchain style solutions. However, in countries with lower social trust scores, the use of blockchain and record-keeping may have a significant impact. We may not fully understand the implications of trust as we embrace the digital society, but we will have to deal with them sooner rather than later.

Bibliography

Adams, C., Cain, P., Pinkas, D. and Zuccherato, R. (2001), 'Rfc 3161: Internet x.509 public key infrastructure time-stamp protocol (tsp)', *IETF Proposed Standard* . https://doi.org/10.17487/RFC3161.

Arkivverket (2018), Noark 5 norsk arkivsystem versjon 5.0, Standard, The National Archives of Norway, Oslo, Norway.

Bakke, F. A. (2017), 'Titalls millioner i overforbruk og prosjektproblemer i politidirektoratet har vært skjult for offentligheten', *Aftenposten* . https://www.aftenposten.no/norge/i/1aEPl/Titalls-millioner-i-overforbruk-og-prosjektproblemer-i-Politidirektoratet-har-vart-skjult-for-offentligheten.

Christina, Q. (2018), 'Kommuneansatte dømt til 6 og 3 1/2 års fengsel for grov korrupsjon', *Verdens Gang* . https://www.vg.no/nyheter/innenriks/i/8w7gRE/kommuneansatte-doemt-til-6-og-3-1-2-aars-fengsel-for-grov-korrupsjon.

Commission, E. (2016), 'Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation)', https://eur-lex.europa.eu/eli/reg/2016/679/2016-05-04.

Crosby, M., Pattanayak, P., Verma, S., Kalyanaraman, V. et al. (2016), 'Blockchain technology: Beyond bitcoin', *Applied Innovation* **2**(6-10), 71.

Duranti, L. (1997), 'The archival bond', *Archives and Museum Informatics* **11**(3-4), 213–218. https://doi.org/10.1023/A:1009025127463.

Duranti, L. and Blanchette, J.-F. (2004), The authenticity of electronic records: the interpares approach, *in* 'Archiving Conference', Vol. 2004, Society for Imaging Science and Technology, pp. 215–220. https://pages.gseis.ucla.edu/faculty/blanchette/papers/ist2.pdf.

Evans, C. (2003), *A question of evidence: the casebook of great forensic controversies, from Napoleon to OJ*, John Wiley & Sons.

Findlay, C. (2017), 'Participatory cultures, trust technologies and decentralisation: Innovation opportunities for recordkeeping', *Archives and Manuscripts* **45**(3), 176–190. https://doi.org/10.1080/01576895.2017.1366864.

Gavin, W. (2015), Poa private chains. https://github.com/ethereum/guide/blob/master/poa.md.

Hagen Sataslåtten, O. (2017), 'The norwegian noark model requirements for edrms in the context of open government and access to governmental information', *Tidsskriftet Arkiv* **8**(2). https://doi.org/10.7577/ta.2485.

Hofman, D., Lemieux, V. L., Joo, A. and Batista, D. A. (2019), '"The margin between the edge of the world and infinite possibility": Blockchain, gdpr and information governance', *Records Management Journal* **29**(1/2), 240–257. https://doi.org/10.1108/RMJ-12-2018-0045.

Høiaas, P. B., Hansen Rørås, C. and Sødring, T. (2016), 'Uttrekkssammenligning - to uttrekk fra samme system', *Tidsskriftet Arkiv* **7**. https://doi.org/10.7577/ta.1670.

ISO-15489-1 (2016), ISO 15489-1:2016 Information and documentation – records management – part 1: Concepts and principles, Standard, International Organization for Standardization, Geneva, CH.

ISO-16175-2 (2011), ISO 16175-2: 2011 Guidelines and functional requirements for records in electronic office environments, Standard, International Organization for Standardization, Geneva, CH.

JEITA (2019), 'Exchangeable image file format for digital still cameras : Exif', https://www.jeita.or.jp/cgi-bin/standard_e/list.cgi?cateid=1&subcateid=4.

Jirgensons, M. and Kapenieks, J. (2018), 'Blockchain and the future of digital learning credential assessment and management', *Journal of Teacher Education for Sustainability* **20**(1), 145–156. https://doi.org/10.2478/jtes-2018-0009.

Kleven, Ø. (2016), Nordmenn på tillitstoppen i europa, Quarterly report, Statistics Norway. https://www.ssb.no/kultur-og-fritid/artikler-og-publikasjoner/nordmenn-pa-tillitstoppen-i-europa.

Kulturdepartementet (2018), 'Forskrift om utfyllende tekniske og arkivfaglige bestemmelser om behandling av offentlige arkiver (riksarkivarens forskrift)', https://lovdata.no/forskrift/2017-12-19-2286/\T1\textsection3-1.

Leach, P., Mealling, M. and Salz, R. (2005), 'Rfc 4122: A universally unique identifier (uuid) urn namespace', *IETF Proposed Standard, July* . https://doi.org/10.17487/RFC4122.

Lemieux, V. L. (2016), 'Trusting records: is blockchain technology the answer?', *Records Management Journal* **26**(2), 110–139. https://doi.org/10.1108/RMJ-12-2015-0042.

Lemieux, V. L. (2017*a*), Blockchain and distributed ledgers as trusted recordkeeping systems, *in* 'Future Technologies Conference (FTC)', Vol. 2017.

Lemieux, V. L. (2017*b*), A typology of blockchain recordkeeping solutions and some reflections on their implications for the future of archival preservation, *in* '2017 IEEE International Conference on Big Data (Big Data)', IEEE, pp. 2271–2278. https://doi.org/10.1109/BigData.2017.8258180.

Lemieux, V. L., Hofman, D., Batista, D. and Joo, A. (2019), 'Blockchain technology & recordkeeping'. http://armaedfoundation.org/wp-content/uploads/2019/06/AIEF-Research-Paper-Blockchain-Technology-Recordkeeping.pdf.

Lemieux, V. L. and Sporny, M. (2017), Preserving the archival bond in distributed ledgers: a data model and syntax, *in* 'Proceedings of the 26th International Conference on World Wide Web Companion', International World Wide Web Conferences Steering Committee, pp. 1437–1443. https://doi.org/10.1145/3041021.3053896.

Litan, A. and Leow, A. (2019), Hype cycle for blockchain technologies, 2019, Report, Gartner, Inc.

Mueller, R. S. (2019), *The Mueller Report: Report on the Investigation into Russian Interference in the 2016 Presidential Election*, WSBLD.

Nakamoto, S. (2008), 'Bitcoin: A peer-to-peer electronic cash system'. `https://bitcoin.org/bitcoin.pdf`.

Soedring, T. (2013), 'Simplechain', `https://gitlab.com/OsloMet-ABI/SimpleChain`. GitHub repository.

Soedring, T. and Reinholdtsen, P. (2013), 'Nikita noark5 core', `https://gitlab.com/OsloMet-ABI/nikita-noark5-core`. GitLab repository.

Werbach, K. (2018), *The Blockchain and the New Architecture of Trust*, MIT Press. `https://doi.org/10.7551/mitpress/11449.001.0001`.