Proof of concept for an XBRL report indexer with integrity and non-repudiation secured by Blockchain using a smart contract: XBRLchain demo

Ignacio Boixo*, Javier Mora**, Jesús Ruiz***

* Founder at Openfiling Association, ignacio@boixo.com
** Manager at XBRL Spain, javier.mora@xbrl.es
*** CTO at Alastria Blockchain Ecosystem, jesus@alastria.io

19 February 2019

Abstract: Financial reports are nowadays being published in the structured format XBRL (Engel, Hamscher, Shuetrim, Kannon, & Wallis, 2003) and academic research has been carried out focusing on the advantages and pending challenges (Perdana, Rodd, & Rodhe, 2014). However, the method of publication itself has received much less attention. Basically, the financial reports are published by the relevant Securities Supervisors on their respective websites. The inherent underlying weakness is that once the financial report has been downloaded from the website, there are no practical ways to maintain certain security characteristics, such as Integrity and Non-repudiation. This article advances the use of Blockchain to preserve the Integrity and Non-repudiation of financial reports along the whole supply chain. Furthermore, the different propagation time of the methods to publish reports on the Internet may provide not illegal insider trading opportunities that Blockchain may be able to mitigate. A proof of concept implementation has been carried out using the operational Alastria Blockchain, which may serve as a benchmark standard between theory and practice. This article and demo explains how to build a distributed ledger platform prototype from the ground up, specifically designed to manage financial reports, in the line of other proposals in the financial area. as Corda for financial agreements (Brown, 2016).

Index Terms: XBRL, Blockchain, RegTech, FinTech, regulatory disclosure, financial reporting, business report, Notarization, integrity, non-repudiation.

Content:

I. INTRODUCTION ............................................................................................................................... 2
II. WHAT IS THE PROBLEM TO SOLVE? .......................................................................................... 2
III. INSIDER TRADING RISK AND POTENTIAL MITIGATION WITH BLOCKCHAIN ...................................................... 5
IV. USE CASE ....................................................................................................................................... 7
V. CONCLUSIONS AND FUTURE WORK ........................................................................................ 9

Appendix 1. Source code .................................................................................................................... 10
Appendix 2. Blockchain technical details .......................................................................................... 10
Copyright .............................................................................................................................................. 10
Who is who? ........................................................................................................................................ 11
Bibliography ......................................................................................................................................... 11
I. INTRODUCTION

The key point in regulatory disclosure is that a business report is legally binding for the issuer when that business report is filed by the issuer to the supervisor and, consequently, published by the supervisor. At that point the Supervisor is providing the characteristic of (Integrity) by publishing the original report with legal validity, as well as the characteristic of (Non-Repudiation) by acting as a Notary of the business report: The issuer cannot deny having sent the business report.

The customary approach to secure a business report is simply to download it from the supervisor’s website, as exemplified in the cases of Spanish Securities Authority (CNMV, 2019). The consumer is therefore confident that the report is the actual report legally filed by the issuer to the supervisor. However, the business report itself lacks any security feature. In the case of a non-supervisory search engine, such as Arelle Filing Index (Fisher, 2019), the solution is to hyperlink by URL to the USA Securities and Exchange Commission (SEC).

What if the issuer simply put its digital signature in the report? From an IT point of view, that is an optimal solution. But, in practical terms, the existing digital signatures depend on Certification Authorities, which unfortunately is far from establishing reciprocal trusted relations, and therefore lacking interoperability and extensibility for most consumers. Moreover, the digital signature must be published somewhere, well concatenated to the report, or in safekeeping in an independent file. Hence, each consumer must download independently the same report from the supervisor website as a security feature, this download being the only proof of integrity (report is unmodified) and non-repudiation (the issuer cannot successfully dispute its filing -hence authorship-) of the report.

II. WHAT IS THE PROBLEM TO SOLVE?

Let’s consider that the Supervisor’s website acts as a Web Notary. But a Supervisory-based Notarization leaves out relevant issuers’ non-regulatory reports; hence, certain reports are not published by the Supervisor, such as Corporate Social Responsibility, Integrated Reporting, Environmental Impact, and so on. This scenario is represented in Diagram 1.

![Diagram 1](image)

Anna and Bob trust the public website of their respective national Supervisor or Officially Appointed Mechanism. However, when Carol receives a report from a third-party supply chain, she has no other method to check its Integrity and Non-repudiation than to compare the report with its original supervisory publication on a website. Note that the URL of the original supervisory publication is not necessarily part of the report. To locate a particular report on a foreign Supervisor’s website is not necessarily an easy task.

Diagram 2 below shows the same situation, but with a public and immutable ledger that cannot be changed, based on Blockchain:
Carol can now check the report’s Hash code (Cryptographic hash function) using whatever Check Server one wants (or directly if one owns an Observer node). If the hash code is found in the Ledger, Carol can be sure of (a) the Integrity of the report (no modifications) and (b) the Identity of who Notarized the report: such Notary, strongly identified by the Blockchain cannot successfully dispute its Notarization: Non-repudiation.

Note that the report must not necessarily be public. As the hash code is an inherent characteristic of each particular report, having the report derive its hash code is trivial. However, having a hash code provides no information at all about the original report. This (One-way function) is known as (Cryptographic hash function). Making a hash code public does not compromise the report’s confidentiality: It is equivalent to getting a key, but ignoring where the lock is.

However, nothing is perfect. When Anna consults the OAM website used by Bob, or when Bob consults the Supervisory website used by Anna, or Carol consults either the OAM website or the Supervisory website, all three people may find difficulties if they are not knowledgeable about the features of each website. This situation is represented in Diagram 3.
This approach is an evolution (currently in Proof of Concept (PoC) development) of the XBRLchain concept, which additionally stores the stable URL of each report in the Blockchain, in addition to the hash code and the identity of the person who has notarized the report. This strategy of linking a URL with the Hash code of the object addressed by such URL is in the roots of the SubResource Integrity (Devdatta, Braun, Marier, & Weinberger, 2016) recommended by the W3C.

A characteristic of Blockchain is that all the nodes have an online local copy of the ledger. Therefore, when someone notarizes a report, such Notarization is simultaneously updated in all the nodes. A smart contract can be implemented, which triggers an event when a new Notarization is generated. A Financial Gateway (currently in PoC development) can immediately read the report from the publication website by using the Notarized URL and including its attributes in an index of reports. See the sequence in Diagram 4:

![Diagram 4](image)

David can now query the updated Financial Gateway for the reports matching a set of attributes and obtain a list of URs. David can then download the selected reports directly from the respective websites. There are other interesting features of this approach which, for the sake of brevity, are not represented in the Diagram.

The Financial Gateway can also (a) check the hash code of the reports, (b) store the reports in full (e.g., as a backup or download accelerator of original publishing websites), (c) safely index non-regulatory reports (e.g., environmental, social responsibility, and so on), (d) be a Business Intelligence public server, and (e) acting as a Notary by default if a publication website fails in its Notarization function, among many other functions.

Moreover, there can be several Financial Gateways running independently, without interferences. Note that no agreement, communication channel, link, RSS, or any special mechanism is required between the Gateway and the publication nodes because the Notarization in Blockchain is enough. All the information for Starting from scratch the index of reports or re-synchronizing it can be obtained by reading the Notarizations in the local copy of the ledger. As the Notarizations are propagated simultaneously to all the Blockchain nodes, and the publication websites are of public access by nature, there are no restrictions for entrepreneurs or public initiatives in building all the Financial Gateways with specific features they want.

This kind of challenge will surely start to surface in Europe in 2020 when all listed companies will be required to report in the same European Single Electronic Format (ESEF) on about 30 different websites, in accordance with the initiative of the European Financial Transparency Gateway (EFTG). Aim of the EFTG Pilot Project consists in developing a Blockchain platform infrastructure technically enabling citizens and investors by giving them increased accessibility to public regulated information provided by the participating Officially Appointed Mechanisms (OAM). The EFTG Pilot Project also aims to provide search capability among real Member States’ data, such as those from the Annual Financial Reports (AFR) in order to contribute to further integrated capital markets. EFTG is based on a new approach to build a distributed and decentralized platform system by interconnecting all the OAMs in Europe into a dedicated platform for sharing data instead of exchanging it, adding also a full traceability and ownership management of the submitted and consumed financial data. Initial references are a video (Szabó, 2017) and presentations (Sel, 2018) (Piechocki, 2018) related to the project EFTG SMART 2016/9488, included in the Financial Technology Action Plan (FinTech-MEMO-18-140612, 2018) of the European Commission.
III. INSIDER TRADING RISK AND POTENTIAL MITIGATION WITH BLOCKCHAIN

On 15 January, 2019, the Securities and Exchange Commission (SEC, 2019) released: SEC Brings Charges in Edgar Hacking Case. The SEC’s complaint alleges that after hacking the newswire services, Ukrainian hacker Oleksandr Ieremenko turned his attention to EDGAR and, using deceptive hacking techniques, gained access in 2016. Ieremenko extracted EDGAR files containing non-public earnings results. The information was passed to individuals who used it to trade in the narrow window between when the files were extracted from SEC systems and when the companies released the information to the public.

The scheme was described (Aitken, 2019) as: [Ieremenko] obtained non-public “test files,” which issuers can elect to submit in advance of making their official filings to help make sure EDGAR will process the filings as intended. It is explained that issuers sometimes elected to include non-public information in test filings, for example actual quarterly earnings results not yet released to the public. And, in Ieremenko’s case he extracted non-public test files from SEC servers, and then communicated such information to different groups of traders.

Insider trading (Chen, 2019) is the buying or selling of a security by someone who has access to material non-public information about the security. Insider trading can be illegal or legal depending on when the insider makes the trade. It is illegal when the material information is still non-public.

In this case, the SEC reported that the non-public information was illegally accessed by deceptive hacking (Barnett, 2011)

Are the current publication systems potentially offering windows for non-illegal insider trading on SEC information?

Let’s review the process for publishing SEC information. A typical listed company (a) publishes its report customarily in the investor relations section of its website (Investopedia, 2019), and (b) submits the report to the SEC, which in turn (c) publishes the report on the EDGAR website. It also (d) provides Really Simple Syndication (RSS) feeds for EDGAR structured disclosure submissions as a courtesy for those interested in viewing, analysing, and manipulating the submissions. These feeds are updated every ten minutes (SEC, 2017).

The above Diagram 5 shows how a (daemon) is waiting constantly for updates on (a) the website of the listed companies and/or (c) the EDGAR system may offer up to ten minutes of non-illegal insider trading window in relation to the investors waiting for (d) delayed RSS feeds. Even one minute is much more than the time needed for high frequency trading (Chen, 2018) or even for any “slower” algorithmic trading (Chen, 2018).

Running a computing process daemons which continuously queries companies and EGDAR websites for updates is frequently used as a technical approach. However, it is not a very elegant solution. In fact, this approach creates a lot of useless traffic between update and update and can eventually overload the web infrastructure.

In the hypothetical case of reports submitted by European listed companies, the situation would be even worse. The European listed company (a) publishes (or not) its report on its own website and (b) submits the report to the national OAM, which in turn (c) publishes the report on the national OAM’s website and may, or may not, (d) provide RSS feeds.

In turn, the OAM (e) would notify the European Financial Transparency Gateway (EFTG), which in turn (f) indexes/republishes the report and/or (g) provides RSS feeds or the equivalent. In short, there are multiple time offsets, hence creating legal insider trading windows, depending on how the investor is notified about the existence of the released report.
By nature, Blockchain offers a very interesting characteristic: all the nodes are simultaneously fed with each new block for updating the local Blockchain copy. Each Blockchain node can explore, in parallel, each new block that detects a new entry of a Notarized Report. To the greatest extent possible, in order to prevent any possibility of non-illegal insider trading, a listed company may opt to (z) Notarize the Hash of the report with the URL of publication (but not actually publishing), and (y) submit the report to the OAM, then (x) wait for the OAM’s acknowledgment and (w) publish the report on the URL. Even when the OAM and/or the EFTG in turn Notarizes again the Hash of report, this is simply the same Hash that has been notarized several times. Diagram 7 shows the sequence.

In step (z) of Notarization, all the Blockchain nodes are simultaneously informed of the Report’s URL. Anna and Bob receive the information simultaneously by nature. There are no possibilities of legal insider trading windows for specific investors. A generalization is show in Diagram 8.

Irrespective if the Report is first notarized by the Issuer, by the OAM or by the Financial Gateway, Anna and Bob will receive simultaneously the notification in the next Blockchain update. Simultaneous update is a Blockchain characteristic (Church, 2017).
IV. USE CASE

XBRLchain is a proof of concept demo using Blockchain to secure the integrity and non-repudiability of a document, which is typically an XBRL instance document containing a business report.

A. How does XBRLchain work? The document is uploaded and assigned a unique and secure code with a standard hash algorithm, and this code is indelibly written (Notarization) by an authorized party in the Blockchain ledger. Checking the validity of a document is done by recalculating the code and comparing it with the one in the ledger, in addition to retrieving the identification of the authorized party. The code provides integrity (the report is unmodified) and the identification of the authorized party provides non-repudiability (the authorized party cannot successfully dispute the Notarization of the report). It is important to note that several authorized parties (issuer, auditor and supervisor) can independently sign the same document. The identification of the issuer, the type of report, the period, etc., is already mandatorily included in the XBRL instance document, in accordance with the XBRL Specification. The only requirement is that the XBRL instance document must be self-containing (i.e., not dependent on external references). Additional information about the authorized party and a timestamp can easily be added.

B. How is a document Notarized? The authorized party requests the services of a notary webserver which is authorized to write to the Blockchain. The Blockchain system must have a secure interface with the server, as well as a smart contract definition interacting with the ledger. See figures 1 and 2.

C. How is a document Checked? Each check web server with reading access to the Blockchain can retrieve and check each hash code generated by the respective authorized party/parties acting as a notary/notaries for such hash code. For demo purposes in this proof of concept, the code is a simple (CRC32) concatenated as an XML comment at the end of the file. See figures 3 and 4.
A live demo of the system is available at the XBRLChain website www.xbrlchain.info where the user can check the following file examples (use userId/Password test/test for Notarize, but only once every 10 minutes). See table 1 below.

<table>
<thead>
<tr>
<th>Filename</th>
<th>TESTING CODE?</th>
<th>Valid hash code?</th>
<th>Registered in Blockchain?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example_1.xhtml</td>
<td>No</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>Example_2.xhtml</td>
<td>Yes (2676830881)</td>
<td>No (must be 2676830882)</td>
<td>N.A.</td>
</tr>
<tr>
<td>Example_3.xhtml</td>
<td>Yes (1751792560)</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Example_4.xhtml</td>
<td>Yes (2676830882)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 1.

This Proof of Concept (PoC) has been created and managed by XBRL Spain on a standard web hosting service. This web page is open to the public, with testing credentials for Notarization and free for Checking. The website is a Client that invokes an Observer Node of Alastria. The development of this PoC is a webpage with two functions: Notarize and Check. See figure 5 below.

For Notarizing, the Observer Node starts a transaction with the Validator Nodes using the Ethereum Quorum protocol. The Validator Nodes prepare a block with the ordered list of transactions (if any) received each cycle of one second (or less depending on workload). The block is distributed to all the Nodes (following the Blockchain architecture), which execute the Smart Contract, confirm the result and update the local Ledger copy accordingly. See Diagram 9 above. At the end of the cycle, a new block has been added to the local Ledger copy in each and every one of the Observer and Validator Nodes.

For Checking, the Observer Node simply returns the required information from its local Ledger copy. Note that the Ethereum Quorum throughput for reading is several orders of magnitude better than for writing. See more at the Introduction to Quorum: Blockchain for the Financial Sector (Price, 2018).

Who provides the notary web server/s and the check web server/s? The market does. It could also be an in-house integration of a web server/local service. In this proof of concept, the testing web servers have been developed by the authors. This project has been presented by the authors in Brussels (Boixo, 2018) and in a webinar (Ruiz, Mora, & Boixo, 2018).

V. CONCLUSIONS AND FUTURE WORK

The above proposed project for the use of XBRLchain as an infrastructure layer for the EFTG does not require any changes to the functionality of the current Supervisory/OAM publication websites. The only new feature is to Notarize each report when it becomes public, and even this function can be requested to be assumed by the EFTG acting as Notary by default.

The next steps will be focused on enlarging the proof of concept by automatically taking feeds from XBRL reports from the SEC and other publishers. A query facility will be added to simplify the retrieval of selected reports. Progress on the development of this project is closely linked to the availability of volunteers, as this is an academic research project with no fundraising.

The authors want to thank the Alastria National Blockchain Ecosystem and the Spanish XBRL Association for providing free IT infrastructure.
APPENDIX 1. SOURCE CODE.

All the source code is publicly available at http://www.openfiling.info/Blockchain/
- ProofOfExistence.sol Smart Contract (Alastria)
- NotarizeServlet.java Notarize (Java XBRL ES)
- CheckServlet.java Check (Java XBRL ES)
- CommonFunctions.java Common (Java XBRL ES)
- poe.py Python interface Java <-> Smart Contract

APPENDIX 2. BLOCKCHAIN TECHNICAL DETAILS.

Each Node belongs to a different permissioned company, running an independent installation of the Blockchain software and storing a local copy of the Ledger. With the use of the Blockchain architecture, the network remains fully operational and reliable even if one third of the Validator Nodes are compromised (i.e., ten or more Validator Nodes are captured simultaneously by an undetected attacker). This distributed network architecture provides a very high resilience without a single point of failure and provides a clear advantage over traditional hierarchical approaches in the financial sector.

Why a permissioned Blockchain?

1. Facilities are accessible and there is an available laboratory for the development of Proof of Concepts for test cases. One of the authors of XBRLchain is the Chief Technical Officer of Alastria, the Blockchain Ecosystem running in Spain for worldwide use.
2. Strong Identity capabilities, based on the work of the Identity Commission: Develop an identity solution inspired by the SSI model and based on uPort. Develop mechanisms that allow for the rapid creation of Alastria identities inter-operating with existing identity mechanisms. (Pastor & Menéndez, 2019)
3. Performance and adequacy. Table 2 below shows a comparison of consensus approaches taken from Making Blockchain Real for Business (Tamayo, 2017, p. 21) according to which a permissioned Blockchain such as Alastria is the optimum solution.

![Comparison of consensus approaches](image)

Table 2.

COPYRIGHT

The design of the XBRLchain concept, the development of the proof of concept, and the publication of the results are the intellectual property of the authors, dated 2018-10-07, who worked mostly during non-business hours and on the weekends. The authors hereby grant a license Creative Commons Attribution 4.0 International License (CC BY 4.0, 2019), hence expressly requesting their attribution, as moral authorship right.
WHO IS WHO?

(Alastria, 2017) is a semi-public, independent, permissioned and neutral Blockchain/DLT network, designed to be in compliance with existing regulations, enabling more than 300 associates to experiment with these technologies in a cooperative environment.

(XBRL-Spain, 2004) is a not-for-profit association organized to spread the knowledge and use of technological standards, focused on eXtensible Business Reporting Language (XBRL)

(Openfiling, 2011) is an open community for the practice and filings of XBRL documents with Open Source programs.

BIBLIOGRAPHY


CC BY 4.0. (2019, 2 15). Attribution 4.0 International. Retrieved from https://creativecommons.org/licenses/by/4.0/


