Implementing A Blockchain Based Learning Analytics Platform

Patrick Ocheja
Graduate School of Informatics, Kyoto University
Kyoto, Japan
ocheja.ileanwa.65s@st.kyoto-u.ac.jp

Brendan Flanagan
Academic Center for Computing and Media Studies, Kyoto University
Kyoto, Japan
flanagan.brendanjohn.4n@kyoto-u.ac.jp

Hiroaki Ogata
Academic Center for Computing and Media Studies, Kyoto University
Kyoto, Japan
hiroaki.ogata@gmail.com

ABSTRACT

Learning records represent the activities of learners on a learning platform. Because learning takes place at different institutions, organizations and/or platforms, it is important to connect learning records belonging to the same learner on these various platforms for a wider spectrum of analytics. With decentralization at the heart of the blockchain technology, we show the implementation of a blockchain based learning analytics platform. By using smart contracts, we enforce restricted access to learner's data and empower learners with more control over their learning records. To ensure that learning records are immutable, we use a hashing strategy to detect changes between earliest version of a learning record and subsequently retrieved versions. Finally, we propose some tests to be carried out and identify some concerns.

KEYWORDS

Learning analytics; blockchain; learning data; smart contracts; learning management systems; Ethereum, learning record store; privacy

1 INTRODUCTION

Learning data reflect the activities performed by learners while learning. From information on a learner's behavior to performance in quizzes and assignments, these data form a reference point for evaluating and improving engagement and performance towards realization of learning goals. With many learning organizations and institutions, the multiplicity of different implementations of learning platforms is inevitable. As such, it becomes necessary to ensure a standard for learning data. Common standards such as Tin Can Experience API [1], IMS Caliper Discovery API [2] have been developed to help reduce the burden of system interoperability. It is on the awareness of these standards that learning data silos otherwise known as Learning Record Stores (LRS) are maintained. These record stores form the backbone for learning analytics.

1.1 Limitations of Learning Analytics Platforms

Despite the availability of reference standards for maintaining learning data on an LRS, it is still difficult to achieve interoperability without some limitations. These problems include:

- Connecting learning histories of a learner on different learning platforms on a single immutable trail.
- Ensuring privacy of learners' records with ease of access control.
- Integrating research and production systems for advancing learning.

1.1.1 Connecting Learning Histories. While learners typically move from one provider's learning platform to another, their learning records are stored distinctly and in a disconnected fashion in separate LRSs. Consequently, each system has to pay the cost of growing learner's data from scratch even for very simple cases. While this might not be a repeated effort for first time learners, it is almost impossible to tell if they are truly first timers or not. This also causes a "cold start" problem in training recommender systems due to unavailability of students' previous learning actions [16]. Proposed systems should allow learners to take their learning data with them in the same way they can take their certificates easily from one institution to another.

1.1.2 Privacy, Security and Access Control. This is another challenge faced when sharing learning records with third parties. Although, learning analytics helps in improving the performance of learners [3] [4], Alan and Kyle [5] in one wide and four narrow questions about conditions for learner's privacy, argue that whatever the gains of learning analytics are, they must be commensurate to respecting learner's privacy and associated rights. The psychological trauma that could result from a single point of privacy compromise can be quite devastating as it is possible to reveal more confidential information from a single point [6]. Proposed systems should ensure prioritization of learner's privacy and learners should be in control of their learning data.

1.1.3 Integrating Research and Production Systems. Availability of learning data for research fosters innovation. In cases where learning data are collected from production and/or research systems, learning analytics researchers are often faced with the heinous task of anonymizing personally identifying information in order to protect privacy of stakeholders and consequently impacting negatively on personalized results [7]. As real-time learning data becomes more desirable for learning analytics research [7], it is crucial to develop new ideas on how to carry out such seamless integration and interoperability of both research and production systems while maintaining privacy of stakeholders involved.

1.2 Blockchain Features as a Solution

This work addresses previously identified limitations of current systems in enhancing learning analytics. We propose solutions to mobility of learner's learning records, distributed consensus in maintaining learning history, privacy and access control mechanisms with prioritized learner's interest and interoperability of different systems (production and research). A blockchain is a distributed database of records or public ledger of all transactions or digital events that have been executed and shared among participants [8]. Below, we identify some of the features of blockchain technology that are key to our proposed solution.

1.3.1 Distributed Consensus and Immutability Features. With its first implementation in Bitcoin [9], blockchain technology is based on a distributed consensus where nodes on the network have access to and keep track of all events that occur on the network. Ledger entries are stored as timestamped, chained immutable blocks. To ensure security and consistency of ledger entries, some nodes on the network offer to add new blocks to the ledger by competing among themselves to solve a computationally intensive puzzle known as the Proof of Work. These nodes are called miners and are rewarded for being the first to provide a correct solution to the Proof of Work. The computing power required for solving this puzzle makes it more difficult to rewrite blocks as such rewrite by dishonest nodes would require resolving associated Proof of Work and acceptance of such solution by honest nodes. These features of blockchain technology provide answers to connecting different learning records from different learning providers with high data consistency.

1.3.2 Smart Contract-based Privacy, Security and Access Control. The blockchain technology has a smart contract feature that facilitates enforcing the terms of agreement between two parties in a contract; in this case, between learners and learning providers or between learning providers. We propose policies that are deployable on the blockchain to control data access and ensure privacy of learner's records and mutual interests of learning providers.

1.3.3 Single Ledger, Multiple Participants. We leverage on the distributed consensus and single ledger-multiple-participants features of the blockchain technology to enhance interoperability of both research and production systems. We propose Learning Blockchain APIs and Datastore Wrappers for ensuring seamless and secure communications between the blockchain and LRSs of learning providers. We suggest potential candidates for enforcing non-intrusive access request and provision for foreign systems.

1.3 Related Work

In fields other than education and learning analytics, there is existing research on applying blockchain technology to non-financial products, such as: medical information [10] and domain name registry [11]. Zyskind et al's work on using blockchain to protect personal data provides insight on achieving privacy preservation on a decentralized network with user control and auditing [12]. While these ideas are fundamental to our discovery of our novel approach, there are many aspects of learning systems that present unique problems that need to be solved, such as: connecting distributed or disconnected learning data, smart contract-based privacy and access control frameworks, and interoperability of different learning systems for both research and production environments. This paper proposes an innovative blockchain based system with important modifications to address the specific needs of education and learning systems.

2 BLOCKCHAIN FOR LEARNING ANALYTICS

2.1 Overview

In figure 1, we propose a paradigm shift from current implementations of learning management systems and platforms to the blockchain technology. Block content represent pointers to learning data with ownership and access policies. Nodes on the peer-to-peer network represent learning providers and learners. Learning activities performed by learners on the learning platforms of learning providers on the network are logged on the blockchain as string representation of queries that can be executed on an external database of learning providers to retrieve such activities. To ensure data consistency and immutability, at block creation time, we execute accompanying queries on the external database and include a cryptographic hash of obtained result as part of the block information. Future response from the execution of this query can be compared to the stored hash and if different, the response is invalid and rejected. We propose herein a secure box for executing these queries against providers' databases with reference to the blockchain network in order to maintain established permissions.
In the next sections, we will discuss further the design of our proposed system and the underlying principles.

2.2 Ethereum Blockchain

It is possible to express real-world processes as states and state transition functions. This code representation of real-world processes on a blockchain loosely defines smart contracts. Although present in bitcoin blockchain, Ethereum (eth) is the first to implement a blockchain with a Turing-complete smart contract programming feature [14]. Being Turing-complete is important because it enables writing programs (especially with loop directives) in fewer instructions with efficient use of space. The concept of smart contract lies at the heart of our proposed design as it makes it feasible to enforce required policies and processes by expressing them as executable codes on the blockchain.
In our implementation, we used the open source version of Ethereum written in Golang – Geth [17]. The Geth boot node is setup as a private network, detached from the live version of Ethereum and other nodes can connect to it using similar network id and genesis block configuration. We also define a custom genesis block configuration with lesser difficulty for easy testing and mining of blocks using CPU/GPU on personal computers.

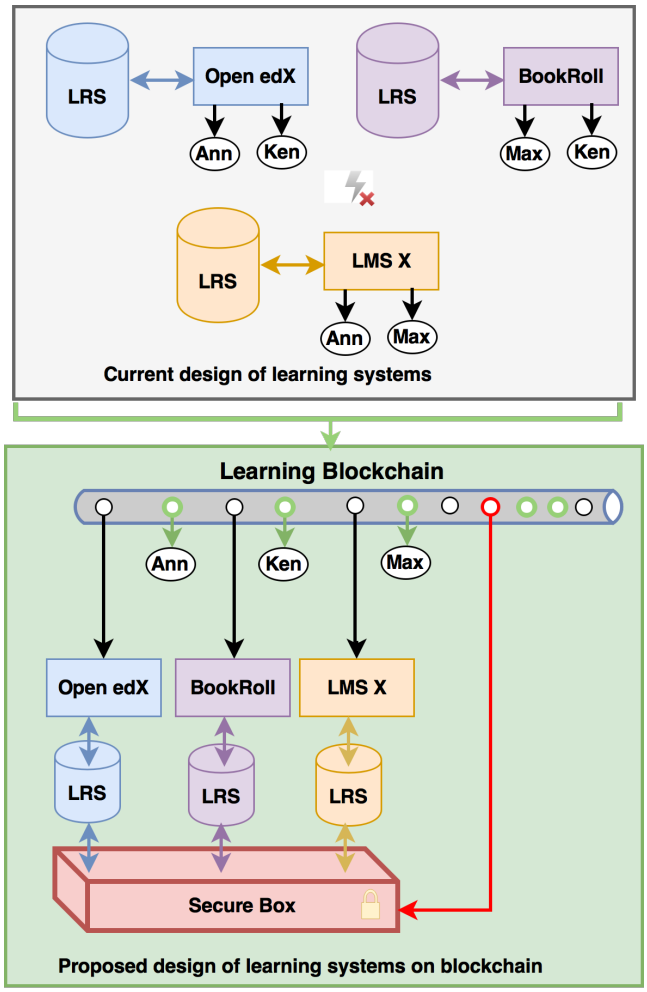

Figure 1. Current learning systems design vs proposed design of learning blockchain

2.3 System Access and Privacy Control

We propose contracts that contain learning data access permissions, ownership and a mapping of the two. The state transition functions of these contracts can be modified to reflect the conditions that must be met before data read or write access is granted. In figure 2, we show the structure of the three main smart contracts namely; Registrar – Learning Provider Contract (RLPC), Learner – Learning Provider Contract (LLPC) and Index Contract (IC) for both Providers and Learners.

2.3.1 Registrar – Learning Provider Contract (RLPC). This contract controls how organizations and institutions become authorized learning providers on the learning blockchain. As these requirements are administratively decided, we propose that typical implementations should consider existing structures for establishing communication and accessing information in institutions and organizations. An example could be the use of special identifiers (ID-1, ID-2, and ID-3 in figure 2) and/or tokens to verify that a node requesting access to the network is actually a known party to the other nodes. This and other conditions can be coded into the RLPC. In our implementation, we keep a record of IP addresses of authorized institutions that can join the network.

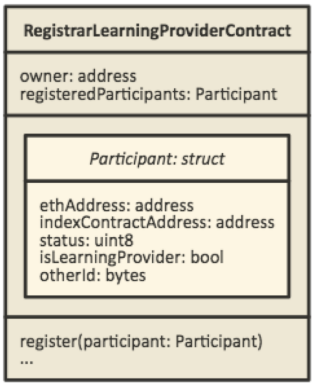

Figure 2. Registrar Learning Provider Contract

2.3.2 Learner – Learning Provider Contract (LLPC). It represents a proof of existence of a learner's learning data on a learning provider's platform. It contains information about the owning learner, address of learning provider's LRS or database with required authentication parameters, queries that can be executed on learning provider's LRS to retrieve learning data, a hash of expected learning data for ensuring data has not been tampered with and a list of access permissions. LLP Contract empower learners with the ability of controlling who can view their learning data by maintaining a list of access permissions granted to other learning providers. The Permission struct in Figure 3 shows an example of some permissions at a very high level.

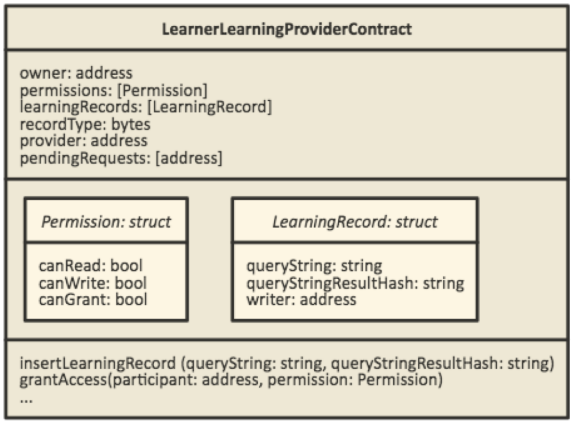

Figure 3. Learner Learning Provider Contract

2.3.3 Index Contracts (IC). An Index Contract contains all LLPCs established between learners and learning providers and by extension, the trail of all learning activities on the blockchain. This is necessary to provide a mechanism for fast lookup of entries and access permissions on the blockchain. We use a hash-table based implementation for the list mapping learners to their LLPCs and another one mapping learning providers to LLPCs they have with learners and with other learning providers that learners have granted access.

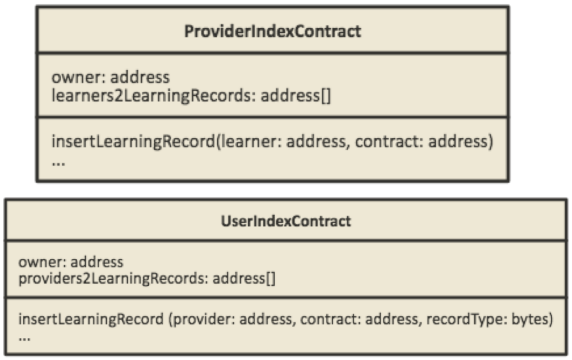

Figure 4. Provider Index Contract and User Index Contract

2.4 Platform Architecture and Process Flow

Figure 5 shows a typical setup of our implementation. We use Moodle LMS[18] and BookRoll [7] as the learning platforms. All learning records emitted on these platforms through learning activities of learners are stored in a central database (MongoDB) through OpenLRW [19]. These learning records are either formatted in xAPI [1] or Caliper standard [2]. We also provide an implementation of a subroutine for retrieving records from the MongoDB through the wrappers on OpenLRW and writing them to the blockchain. On the blockchain, learning records are uniquely grouped using the action verb field and the user's blockchain address. However, before user's or learner's learning records can be written to the blockchain, they must go through the account creation phase.

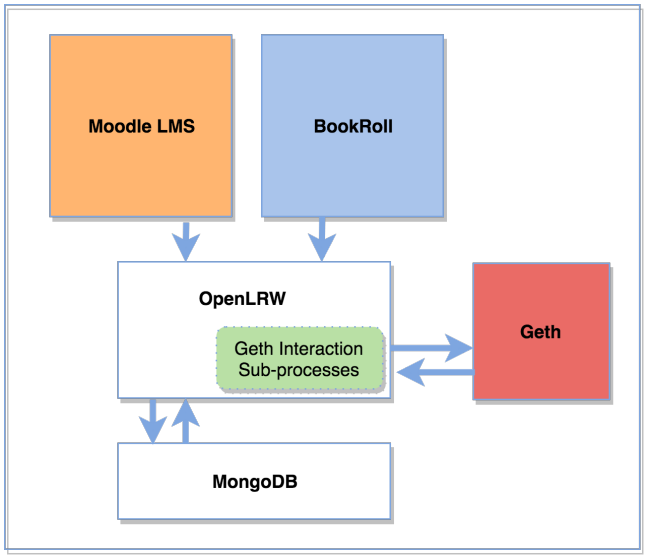

Figure 5. System Architecture – one Institution

2.4.1 Blockchain Account Setup. Learners that opt to have their learning records on the blockchain will have to go through the account setup process. This process handles the generation of blockchain address for the learner, creation of an Index Contract – User Index Contract and the final phase of registering the generated blockchain address and User Index Contract address in the Registrar Learning Provider Contract. Figure 6 shows a flowchart for this process. The process highlighted in red are blockchain dependent transactions that are only completed when the transaction is successfully mined. In a typical scenario, this could range from 30 seconds to several minutes before mining occurs. It is important to note that the only process that requires any action on the part of the learner is consenting to having their learning records on the blockchain and provision of a passphrase to encrypt their private key. The other processes shown in the flowchart are background operations and the decision blocks are solvable by querying the blockchain.

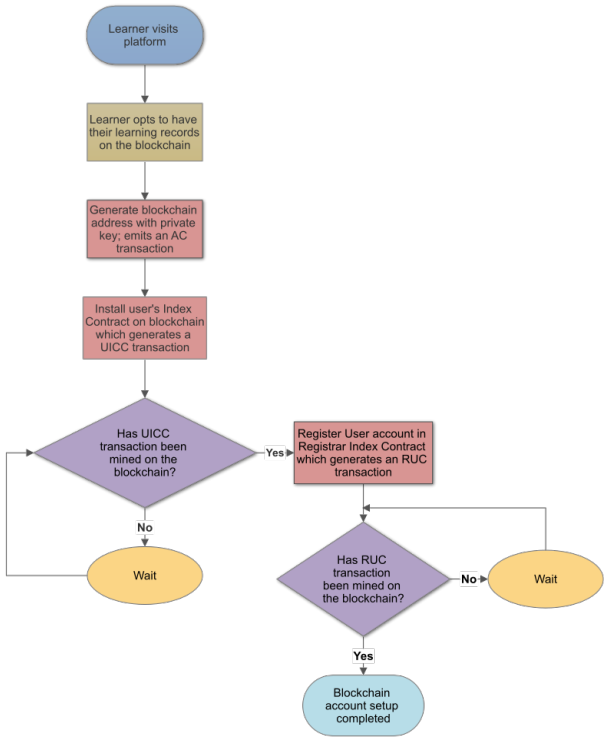

Figure 6. Process flowchart for blockchain account setup

2.4.2 Writing Learning Records. This entails performing at least one transaction on the blockchain. The process begins with retrieving the action verb of the learning record and converting it to a corresponding hexadecimal number. This is required because we want to optimize gas usage on the blockchain; writing strings of variable length require more computational resources in solving the Proof of Work especially when the string is lengthy. After converting the action verb to hexadecimal equivalent, we then query the blockchain to know if a smart contract based on this action verb exists for this user. If it does, we retrieve the smart contract and simply update it with the current learning record's query string and query result hash. If no such smart contract exists for this action verb, we create the smart contract and update the index contracts of both the provider and the learner. The latter case will require four transactions which must be mined on the blockchain. Figure 7 shows the flowchart for this procedure. No user input is required.

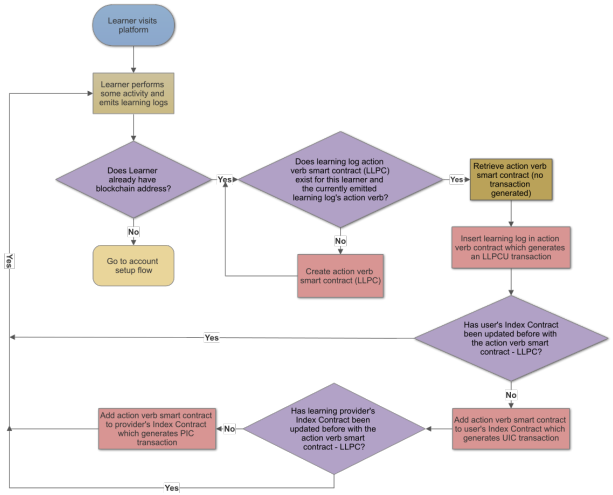

Figure 7. Provider Index Contract and User Index Contract

2.4.3 Accessing Learning Records. In figure 8, we show sequence of activities that occur on the blockchain and how they are handled. At S_0, the blockchain contains only the boot node, RLPC and a Secure Box. KU node then attempts to join the network which prompts verification with established rules in RLPC. Upon successful verification, KU is added as a valid participant and an IC is generated. Learner A (L-A) visits KU's platform and since it is his/her first visit to any node on the network, a new account is created for L-A at S_2. Subsequent learning activities leading to generation of learning data are logged on the blockchain as LLPC-A_n.

At S_{n+1}, edX attempts to read the learning data (LLPC-A_1) of L-A, this is outrightly rejected as there is no proof of edX being aware of the existence of L-A. Later on, L-A decides to visit edX platform and provides their blockchain information to edX. Now, edX knows of the existence of L-A. This means that further request to access L-A's learning data will be forwarded to L-A for approval. If approved, the permission is written on the LLPC and access to the learning data is granted.

3 DISCUSSION

From this implementation, we observe that while some of the transactions on the blockchain require very minimal resources (such as the blockchain address issuing transaction), others require some amount of time; typically, about 2 minutes (tests currently ongoing). For an on-demand connection of learning records, this time might be acceptable. But for real-time connection of learning records it might pose a challenge. Also, we observe that the more the available mining nodes on the network, the faster the transactions are processed (test results to be provided). This ideally follows the tenets of a decentralized network where the best throughput is achieved if everyone mined their own transactions. While it might be difficult to achieve a system where all learners mine their own learning records, it will be interesting to consider alternative approaches to improving on transaction processing time by leveraging on client-side-browser-based mining nodes.

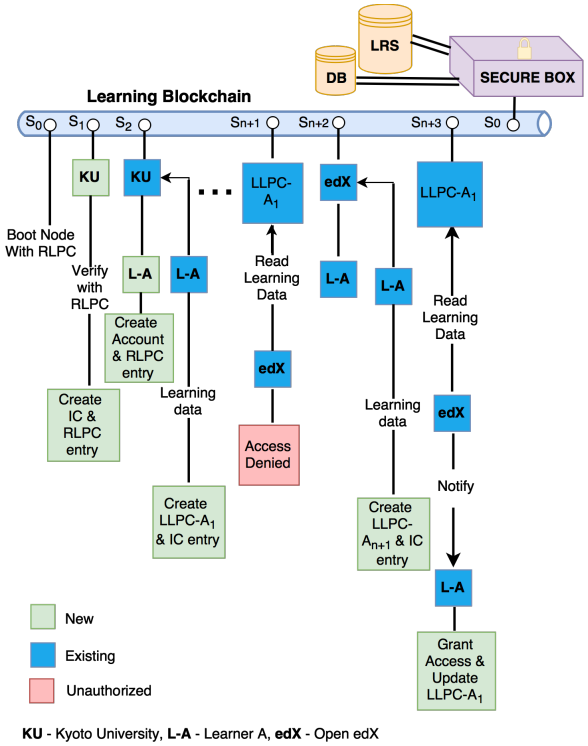

Figure 8. Sample process of registering and accessing blockchain information.

However, one very important concern is the sustainability of the blockchain technology due to the large computing and energy resource requirements. We are aware of this limitation and hence, we have constrained current implementation to use the institution's central resource. Future work should consider possible optimizations to the underlying blockchain technology.

Another important aspect to be considered is defining and enforcing existing user data privacy policies on the learning records using smart contracts. While our implementation considers very top-level approach of representing these permissions, it will be necessary to understand the implications of having 'action verb-based' privacy definitions. Also, we also propose that a further research should be done on how learners can write their own smart contracts using familiar concepts and enforce them on their learning records.

4 CONCLUSION AND FUTURE WORK

In this work, we introduced the concept of connecting learning records using the blockchain. We provided some core aspects of our current implementation which is still been developed. In our future work, we will provide more concrete results on resource requirement, throughput, and a close comparison to alternative systems currently been used.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 16H06304.

REFERENCES

[1] Advanced Distributed Learning. (2016). Experience API (xAPI) Specification. Retrieved from http://github.com/adlnet/xAPI-Spec/
[2] IMS Global Learning Consortium. (2015). Caliper Analytics. Retrieved from http://www.imsglobal.org/activity/caliper
[3] Fumiya Okubo, Takayoshi Yamashita, Atsushi Shimada, Hiroaki Ogata, A Neural Network Approach for Students' Performance Prediction, LAK 2017, pp.598-599, 2017.3.
[4] Sclater, N., Peasgood, A., & Mullan, J. (2016). Learning analytics in higher education. JISC. Retrieved from http://repository.jisc.ac.uk/6560/1/learning-analytics_and_student_success.pdf
[5] Alan Rubel and Kyle M. L. Jones. 2016. Student privacy in learning analytics: An information ethics perspective. The Information Society. Vol. 32(2). 143-159. DOI: http://dx.doi.org/10.1080/01972243.2016.1130502
[6] Omer Tene and Jules Polonetsky, Big Data for All: Privacy and User Control in the Age of Analytics, 11 Nw. J. Tech. & Intell. Prop. 239 (2013). Retrieved from http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1
[7] Brendan Flanagan and Hiroaki Ogata. 2017. Integration of Learning Analytics Research and Production Systems While Protecting Privacy. Chen, W. et al. (Eds.) (2017). *Proceedings of the 25th International Conference on Computers in Education*. New Zealand: Asia Pacific Society for Computers in Education. (in press)
[8] M. Crosby et al. 2015. Blockchain Technology; Beyond Bitcoin, Sutardja Center for Entrepreneurship & Technology. Berkeley Engineering. Retrieved from http://scet.berkeley.edu/wp-content/uploads/BlockchainPaper.pdf
[9] S. Nakamoto. 2008. Bitcoin: A Peer-to-Peer Electronic Cash System. White Paper.
[10] A. Azaria, A. Ekblaw et al., MedRec: Using blockchain for medical data access and permission management, In 2016 2nd International Conference on Open and Big Data (OBD). Institute of Electrical and Electronics Engineers (IEEE), Aug. 2016.
[11] H. Kalodner et al. An empirical study of Namecoin and lessons for decentralized namespace design. WEIS '15: Proceedings of the 14th Workshop on the Economics of Information Security, June 2015.
[12] G. Zyskind et al. Decentralizing privacy: Using blockchain to protect personal data. Security and Privacy Workshops (SPW) 2015 IEEE. IEEE pp. 180-184 2015.
[13] Sony Global Education. 2017. Sony Develops System for Authentication, Sharing, and Rights Management Blockchain Technology. News Release. Retrieved from https://www.sony.net/SonyInfo/News/Press/201708/17-071E/index.html
[14] Buterin, V. 2013. Ethereum White Paper. Retrieved from https://github.com/ethereum/wiki/wiki/White-Paper
[15] SHA-2 Standard. Secure Hash Standard FIPS PUB 180-4. Retrieved from https://csrc.nist.gov/publications/detail/fips/180/4/final
[16] Barnes, T., and Stamper, J. 2008. Toward automatic hint generation for logic proof tutoring using historical student data. In *International Conference on Intelligent Tutoring Systems* (pp. 373-382). Springer, Berlin, Heidelberg.
[17] Ethereum in Go Language. https://github.com/ethereum/go-ethereum
[18] Moodle Learning Management System. https://github.com/moodle/moodle
[19] Open Learning Record Warehouse. https://github.com/Apereo-Learning-Analytics-Initiative/OpenLRW