Doctoral Dissertation

Blockchain of Learning Logs (BOLL): Connecting distributed educational data across multiple systems

Patrick OCHEJA
Supervisor: Hiroaki OGATA

September 2022

Department of Social Informatics
Graduate School of Informatics
Kyoto University

Abstract

Technology and data have continued to play vital roles in redefining various aspects of our lives including teaching and learning. However, learning data continuity is still lacking and the absence of prior learning data creates a *cold-start problem* as the learning data collected at their previous schools are not available for use at the learner's current or future schools. Challenges to enabling learning data continuity include concerns such as security, privacy, interoperability, and lack of enabling infrastructure for analysis of such distributed data. The advent of decentralized technologies such as the blockchain presents a unique opportunity to solve these problems and enable learning institutions and platforms connect the data of their students across multiple environments in a trusted, secure, tamper-proof and traceable way.

In this research, we make three (3) broad contributions. First, we proposed and implemented a Blockchain of Learning Logs (Boll) system: a decentralized system that enables learners to connect their learning statements or logs across different schools they have attended. Second, we proposed and implemented a framework to enable students to access their digital textbooks and learning materials after they change school or graduate. Third, to improve stakeholder awareness and usability of the proposed Boll system, we designed and implemented visualizations for education blockchain data.

Across these three (3) contributions, this thesis investigated the performance of the proposed Boll system, the need, and relevance of prior learning records, and decentralized learning analytics. One system experiment was conducted to investigate the performance of the Boll system and the blockchain when deployed in an education setting. One exploratory study was performed to explore teachers' needs when accessing their students' prior learning data. Two evaluation studies were designed and carried out in K-12 educational setting to investigate the relevance of prior learning data connected on Boll.

In summary, this thesis conducted a theoretical and practical investigation on connecting distributed learning data and analytics with studies on needs, designs and evaluations. The findings suggest that connecting learning data of learners across different schools can be beneficial to both teaching and learning, solves the cold-start problem and further enables lifelong learning and analytics. This research also provides concrete support for enabling personalized learning at scale and enables cross-border analytics of lifelong learning.

Contents

List of Figures

List of Tables

Acknowledgement

First and foremost I would like to sincerely thank my supervisor, Prof. Hiroaki Ogata, for all his support over the past 5 years, from my time as a research student up to the doctor course. During this period, he has not only imparted knowledge on topics within my research area, but also taught me how to conduct and advance research, from conception to fruition, cultivate a deep understanding of the field and how to become an exceptional researcher. His wise guidance and warm encouragement have also enabled me to pursue challenges that I would have only dreamed of in the past.

I am also grateful to my advisers: Prof. Masatoshi Yoshikawa and Prof. Naoyuki Iwashita for their critical assessment and insightful comments, which challenged me as a researcher to achieve a higher standard of work. They have been selfless in their assistance and guidance throughout my research: always available and ready to make time for meetings to discuss my progress and next steps. I have also learnt from them the need to take research beyond a single domain and consider how other fields can benefit from the same research work. Thank you for the opportunity to learn from your wealth of experience and knowledge.

I would also like to thank Dr. Brendan Flanagan for his expert advice on my research and for always being available to review my drafts and provide insightful comments on how to make sound and outstanding arguments. I can not forget how you always helped me to see the good side of every unsuccessful paper submissions. At times when I was at my lowest, you provided the right guidance to set me back on track. Thank you for helping me to grow as a young researcher: I am proud of what we have been able to achieve together. To all the members of Ogata Laboratory and other external collaborators, thank you for your constructive criticism and feedback. I am glad that I had the opportunity to do research in the midst of such a dynamic group of excellent minds.

Finally, I would like to express my deepest thanks to my wife, Gina, for being a strong support. I could not have done this without you. I am also thankful to my family for their love and encouragement. I am especially thankful to my Dad for his ever stimulating words of encouragement. Rest-on Dad. Thank you Mum for everything you have taught me, continue to rest in the bosom of the Lord. To my friends Tunny, Paschal, Uche, Abu, Kunle, Chenemi, Grace, Ibrahima, Wilson and everyone who has been helpful (indeed, it takes a whole village to raise a child): thank you for being there for me.

Chapter 1

Introduction

Technology and data have continued to play vital roles in redefining various aspects of our lives including teaching and learning. While technology can make learning more readily available, accessible and multimodal, the data collected on learning platforms avail useful insights to help learners achieve their goals (Chatti et al., 2012). However, learning data continuity is still lacking and the absence of prior learning data creates a *cold-start problem* as the learning data collected at their previous schools are not available for use at the learner's current or future schools(Barnes & Stamper, 2008). Challenges to enabling learning data continuity include concerns such as security, privacy, interoperability, and lack of enabling infrastructure for analysis of such distributed data (Baker et al., 2019).

The advent of decentralized technologies such as the blockchain (Nakamoto, 2008) presents a unique opportunity to solve these problems and enable learning institutions and platforms connect the data of their students across multiple environments in a trusted, secure, tamper-proof and traceable way. But such implementation is still lacking due to absence of supporting frameworks, tools, interoperability specifications and certain affordances peculiar to the education sector. More so, there has been limited research on blockchain in education with concrete implementation and evaluations on usefulness to teaching and learning, usability, feasibility, resource requirements, and limitations (Ocheja et al., 2022).

In this research, we proposed and implemented a Blockchain of Learning Logs (Boll) system, conducted and reported experiments regarding its use and relevance to teaching and learning. The Boll system is a decentralized system that enables learners to connect their learning statements or logs and digital contents across different schools they have attended. The Boll system enable various learning institutions to securely interact with one another in a decentralized manner using a public and verifiable ledger. The following section present key problems addressed by this thesis including transferability, privacy and security, verification, consistency and traceability, and distributed analytics of learning logs and access to digital contents.

1.1 Problems

1. **Connecting learning logs across multiple institutions**: Learning logs of learners exist across multiple institutions due to the lack of connection among their LRS's. Take for instance a student who studied at n different institutions becomes responsible for managing their learning logs in n different places. Also, consider that each of these n-institution cannot seamlessly access the student's data across their systems. Consequently, each system has to acquire the learner's data afresh even for very simple cases. While this might not be a repeated effort in the case of first time learners, it is almost impossible to tell if these learners have previously interacted with other learning systems or not. This scenario of disconnected learning logs is shown in figure 1.1. As these students change school, they leave behind their past learning logs and only move to the next school with their certificate and/or transcript. To solve this problem, it is necessary to provide a mechanism for allowing interaction between learning systems such that learning logs of students can be connected across these systems.

Table 1.1: Comparison of the different database architectures

Property	**Centralized Databases**	**Distributed Databases**	**P2P Decentralized Databases**
Storage	Centralized	Replicated in multiple places	Replicated by all participants
Access authorization	Issued by a central node	Issued by distributed nodes	Issued by actual data owner
Modification	Could be tampered	Could be tampered	Impossible to modify maliciously
Write Efficiency	High	High	Low
History trace	No	No	Yes
Trust	Single authority	Few nodes	By consensus

One approach to connecting learning logs could be by providing a central database for all learning logs from these institutions. But as shown in table 1.1, centralized databases

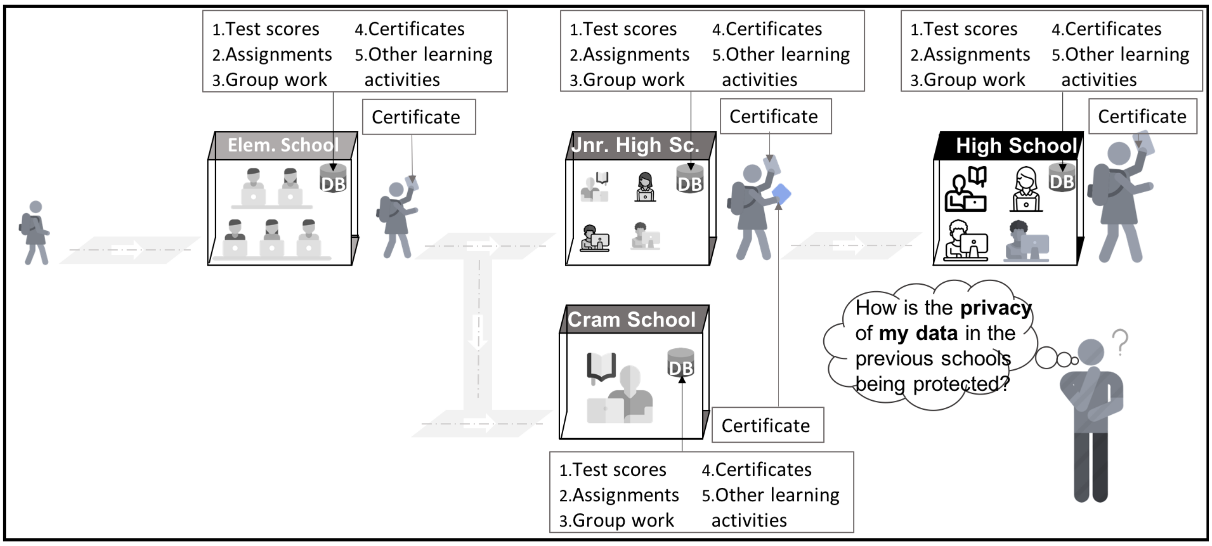


Figure 1.1: Disconnected learning logs.

have some limitations such as the inherent centralization and control by a single authority such as a country's department or ministry of education. The implications of such a central control could range from ease of tampering with records to inability to ensure compliance with privacy regulations. Also, distributed databases may overcome the single-point of failure problem with centralized databases but distributed databases do not provide a mechanism for enabling direct control of records by actual data owners. It is often common to refer to distributed databases as decentralized databases but in this thesis, decentralized databases refer to databases whose transactions are driven by a peer-to-peer (P2P) network as distinguished in (Bonifati et al., 2008). Decentralized databases like the blockchain, a P2P network, provide a mechanism for enabling trust by consensus where all parties can decide on the rules and transparently audit the activities of other parties on the network. Also, decentralized databases provide an inherent mechanism for accessing transaction history and ensure that processed transactions cannot be altered by malicious actors. This thesis proposes the use of the decentralized architecture of the blockchain to facilitate interaction between multiple institutes towards connecting the learning logs of their students.

2. **Enable transfer of learning logs**: Apart from transcripts and/or certificates, learn-

ers are unable to transfer their past learning records to their current institution. Connecting learning records at different institutions provides a student with a journal of lifelong learning but allowing transferability of learning records allow these different institutions to access the student's data in order to provide a personalized learning experience through learning analytics or provide a data-driven education. When learning logs are non-transferrable, conducting research using learner's data becomes limited. More so, when researchers do not have access to real-time learning logs, advancing learning through interventions by detecting at risk students becomes difficult. As real-time learning data becomes more desirable for learning analytics research (Flanagan & Ogata, 2017a), 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 the privacy of stakeholders involved.

3. **Enable transfer of learning materials**: Similar to learning logs, learners are unable to transfer their past learning materials across schools. For example, a student who attended an undegraduate program in Computer Science (CS) at School A and enroled in a course on Data Structures and Algorithms. In this course, they may have used different digital textbooks and lecture slides provided through their school's e-learning system. When the student graduates and enrols in a graduate school program in CS, revisions on Data Structures and Algorithms might be crucial to their success on certain topics. However, because they have graduated, it is impossible for them to access the same learning resources. Thus, the learner may have to source for other literature to use for their revision. Teachers also find this situation problematic as they are unable to tell beforehand what learning resources their students might have used in the past or what teaching methods they are more accustomed with. More so, lack of transferability of learning materials means that teachers at the student's new school are unable to know the depth of learning, for example, what type of assessment questions the student previously solved on prerequisite topics and what areas might be perturbing.

While the above scenario applies to students changing school, another situation where decentralized access to learning materials is desirable is in a situation where a sponsor pays a publisher to grant students access to specific learning resources. In a typical education context, the sponsor could be the ministry of education paying for K-12 resources. In this context, the sponsor wants to pay for only the resources the students have used and the publisher would also want to receive the exact payment as well. When the learning resources are hosted on the school's learning infrastructure or the publisher's publishing platform, both parties would require some level of trust to agree on usage data and corresponding payment. Hence, for potentially distrustful parties, it becomes necessary to facilitate such contracts in a decentralized and transparent manner. Also, the resulting data from students' use of such learning materials could be useful to both the publisher and learning institutions, thus, requiring the permission of relevant stakeholders to grant such access. The ability to transfer learning logs accompanied by the respective learning materials makes possible for a robust learning analytics and comprehensive journal of lifelong learning.

In this thesis, we proposed a framework that allow learners transfer their learning materials such as digital textbooks, lecture slides and assessment questions/solutions across schools. We design smart contracts that can be used by authors to protect their work and offer various license types to learners who want to use their published works. Learners can in turn review these contracts and decide whether to subscribe or not.

4. **Facilitating privacy of learning logs**: The lack of protection and control of private information by data owners exist as a result of the disconnection between different LRS's. An example to explain this problem is how students move from one school to another but become less aware of how their past learning logs are being used. In figure 1.1, the learner at end of their learning begins to wonder if the privacy of their learning logs at their past schools is being respected or not. Although, learning analytics helps in improving the performance of learners (Okubo et al., 2017; Sclater et al., 2016), (Rubel & Jones, 2016) used a set of questions grouped into one wide question and four narrow questions to determine the conditions for learner's privacy. They argued that whatever the gains of learning analytics are, such gains must be commensurate to respecting learner's privacy and associated rights. Furthermore, 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 (Tene & Polonetsky, 2012). While connecting learning logs across different systems and engendering transfer of these logs, it is necessary to prioritize learner's privacy: learners should be constantly aware and have control of their learning logs.

A key aspect in learning analytics is the control of personal and private information by

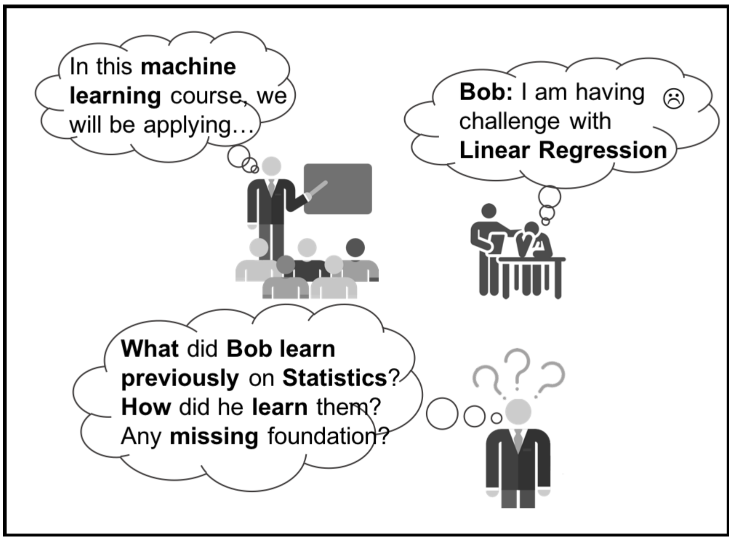


Figure 1.2: Tracing a learner's learning path.

an individual. This includes the ability to opt out of learning activity tracking and giving parents of underaged learners the right to manage their dependents' learning records (36, 2016; Pardo & Siemens, 2014; Rubel & Jones, 2016). Usage or access to a learner's learning records should be sought from the learner and/or their institution depending on the terms of agreement between both parties or according to other defined policies. This agreement should contain clauses such as: usage policies, access authorization, storage policies, etc. The proposed solution in this thesis is to facilitate these agreements on the blockchain by allowing the learner and their institutions to act as signatories on defined smart contracts, and enable the protection of learning records on the blockchain.

5. **Enable traceability of learning**: As learners move from one institute to another, it becomes necessary to know which institutes they have been to previously. One reason for such requirement is a case where a teacher needs to trace the root cause of a particular difficulty experienced by their student. In figure 1.2, *Bob's* teacher is faced with the task of detecting the gap in *Bob's* past learning in a prerequisite course. To detect this gap, *Bob's* teacher needs to know what topics in *Statistics* were covered at *Bob's* previous school and what *Bob's* performance was in each of these topics. Another reason to enable traceability of learning is the need to verify what a learner actually knows. The current way of verifying a learner's knowledge is through tests and examinations. However, this approach suffers from the limitation of not being exhaustive. Connecting learning records on the blockchain provides an additional benefit of enabling traceability. This can be

achieved using the nested transactions feature of the blockchain where the current block contains a reference to the previous block.

6. **Ensure consistency of learning history**: To verify that a learner has previously engaged in a particular learning activity, it is required to ensure that past learning records are unmodifiable. While regular databases provide means for storage of learning logs, they are found to be unable to guarantee that such learning logs have not been altered as identified in table 1.1. This thesis proposes the use of the blockchain to ensure that once learning logs are written, such logs can no longer be altered. The implication of using the blockchain is that if one desires to alter the contents of a past learning log, all institutions on the blockchain network must agree to such an alteration of records.

7. **Reducing the administrative burden of managing multiple institutions**: One common challenge with facilitating interaction among multiple parties is the need to coordinate and ensure that all parties adhere to certain rules. In the education sector, it is common to have a hierarchy where institutes are grouped into different categories using attributes such as location (districts, states, regions, national, etc.), government affiliation (national, public, private, etc.), or core focus (medical school, engineering school, business school, etc.). Managing these different categories would require setting up administrative units to facilitate operations and ensure compliance. In a case where these schools want to share information of their students using traditional means such as distributed databases or proprietary endpoints, more manpower will be required frequently to carryout procedures such as compliance checking. In this thesis, we propose an approach where these administrative tasks can be automated and compliance-check can be conducted using a consensus algorithm. By using a consensus algorithm, only compliant parties can access or write information on the network.

8. **Enabling distributed learning analytics**: It is not just enough to enable learning data continuity: successful and meaningful use of the transferred data is also important. In fact, prior literature have revealed that despite the availability of multisource learner data, interoperability towards meaningful analytics still remains a challenge (Baker et al., 2019). This problem could be caused by factors such as: the difference in learning tools and data standards, lack of interoperability and consistency in semantics, the difficulty in facilitating communication between systems due to privacy limitations as well as other

ethical concerns.

A typical example where distributed learning analytics is desirable is when high school students are at a cross-road on which university program to enrol in. In the traditional context, students would seek professional advice which is not often readily available (van Klaveren et al., 2019), (Stinebrickner & Stinebrickner, 2014). To support such student through learning analytics, it becomes necessary to know what performance level and scores are prevalent among previous applicants to different universities. When we can connect lifelong learning logs across multiple through the blockchain, it is possible to extend such a system to include decentralized analytics of the connected data. Through the permissions feature on a decentralized network, students can opt-in to such services and provide their data for analytics in return for getting insights from other's data.

In this thesis, we propose and implement a framework that enables analysis of distributed data on the blockchain. Through installed smart contracts, learners can choose to opt-in to learning analytics services on the blockchain and receive useful insights such as suitable career path. Our proposed implementation ensures that the privacy rules specified by data owners are not violated.

1.2 Contributions

In this research, we make three (3) main contributions:

1. Connect distributed educational data of learners across different schools: protected, tamper-proof, transferable, verifiable, traceable, and consistent.

2. Enable access to learning resources and usage information across different schools: protected, transferable, reusable and audited.

3. Evaluate the usability, usefulness and impact of connecting lifelong learning data.

Connect educational data of learners across different schools

First, we proposed and implemented a Blockchain of Learning Logs (Boll) system: a decentralized system that enables learners to connect their learning statements or logs across different schools they have attended. The Boll system enable various learning institutions to securely interact with one another in a decentralized manner using a public

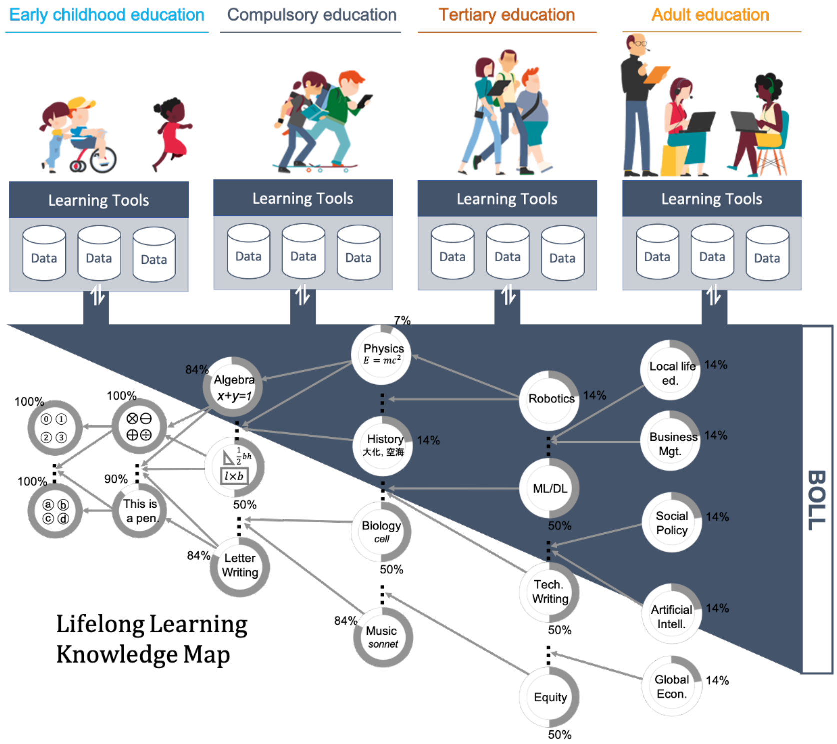


Figure 1.3: Connected lifelong learning: Blockchain of Learning Logs (BOLL).

and verifiable ledger. We present Boll as a solution to the problem of transferring educational data between different institutions as students move from one institution to another. It also solves the cold-start problem in learning analytics systems where a new students' learning environment is created without being informed by previous learning activities, even though their current learning activity is based on experiences at their previous school. With Boll, previous learning data could serve as a robust foundation upon which new learning environments are created when a learner enrolls in a new institution. On the proposed blockchain network, learning institutions can coexist, exchange information and maintain reference to same student's information across multiple systems through a connected trail as shown in figure 1.3.

Enable access to learning resources and usage information across different schools

Second, we proposed and implemented a framework to enable students to access their digital textbooks and learning materials after they change school or graduate. We achieve this goal by extending the Boll system to include a decentralized e-learning marketplace

(Boll-M) where authors, publishers and sponsors can provide students with the required learning materials based on a smart contract policy. Access to prior learning resources can avail learners opportunities to revise and reflect previously taught concepts in preparation to learn a new and related concept. It also provides teachers with a more in-depth diagnostic tool when probing the extent of students' learnings or evaluations on prerequisites. We present Boll-M as a solution to previous barriers to transfer of these digital contents such as intellectual property rights' protection, lack of distributed publishing/access platform, and managing royalties and rewards on usage of intellectual property.The Boll-M features primarily engender trust among sponsors, authors and users of published work by providing a transparent auditing of access logs over a decentralized network.

Evaluate the usability, usefulness and impact of connecting lifelong learning data

Third, to improve stakeholder awareness and usability of the proposed Boll system, we designed and implemented visualizations for education blockchain data. Through the use of learning analytics on education records, and distributed access to these records, we demonstrate how different stakeholders in education can manage, and make sense of their past academic records or that of their students to support different goals. Also, we developed a new big query interface for decentralized learning analytics which can be used to query learning records connected on the Boll system. Our main goal of providing such an interface is to make the learning analytics task of researchers and other stakeholders easier when interacting with a blockchain-based tool like Boll and eliminate the need for blockchain expertise.

Across these three (3) contributions, this thesis investigated the performance of the proposed Boll system, the need, and relevance of prior learning records, and decentralized learning analytics. One system experiment was conducted to investigate the performance of the Boll system and the blockchain when deployed in an education setting. One exploratory study was performed to explore teachers' needs when accessing their students' prior learning data. Two evaluation studies were designed and carried out in K-12 educational setting to investigate the relevance of prior learning data connected on Boll.

The results from the system experiment showed that it is possible to connect learning records of learners across different schools on the blockchain and also revealed key considerations when designing the required smart contracts and on-boarding learners. The exploratory study exposed teacher's lack of access to prior learning data of their students even when they deemed such data as important and conveyed the teachers' needs. The first evaluation study on relevance of students learning data in a prerequisite course at a previous school to current learning at the new school revealed that a significant correlation exist between the engagement levels and scores of highly engaging students and students with very low engagement. The second evaluation study on relevance of access to decentralized learning data as evaluated through higher education enrolment prediction for high school students, showed that students can make better enrolment decisions using decentralized analytics.

In summary, this thesis conducted a theoretical and practical investigation on connecting distributed learning data and analytics with studies on needs, designs and evaluations. The findings suggest that connecting learning data of learners across different schools can be beneficial to both teaching and learning, solves the cold-start problem and further enables lifelong learning and analytics. This research provides concrete support for enabling personalized learning at scale. It will also enable cross-border analytics of lifelong learning. This work can create a paradigm shift in data-driven education to a decentralized approach where all institutes can work collectively to impact knowledge on the learner. The findings have implications for researchers in the learning analytics domain of education research as it presents new methods for multisource data collection and analysis. These findings also have a ripple effect for other domains such as knowledge management, healthcare, AI ethics, design of intelligent and agent-based systems where similar settings of user data continuity, the need for privacy, control, data transfer and analytics exists.

1.3 Research questions

This thesis focuses on answering the following research questions related to the design and evaluation of impact of connecting lifelong learning and analytics:

1. **Design questions:**
    1.1. How to connect distributed lifelong learning logs of students across different schools?
    1.2. How to enable the transfer of digital learning materials across different schools with intellectual property protection and transparent use?

1.3. What mechanisms can we use to enable exchange of information and learning analytics across schools?

2. **Impact evaluation questions:**

   2.1. What cold-start problems does connecting lifelong learning logs solve?

   2.2. What are the perceptions of teachers about prior learning data, lack of and how do they use such information?

   2.3. What is the relevance of distributed lifelong learning data and analytics to learners' future goals?

In chapter 1, we introduced the key problems addressed by this thesis and why it is important to solve them. This include problems such as lack of learning data continuity, non-transferable learning materials and how to enable decentralized analytics. We also highlighted the research questions, our key contributions and novelty different from existing systems.

In chapter 2, we conducted a literature review on prior researches and systems that are related to our work and their limitations. Specifically, we reviewed related learning technologies, decentralized systems and blockchain technology in education as well as distributed analytics. Various gaps were identified and reported in this chapter.

In chapter 3, we presented our solutions to enabling learning data continuity, transfer of digital learning materials and a platform that can enable decentralized analytics. We also discussed the attributes of our proposed frameworks, system architectures and designs.

Chapter 4 contains a detailed report on various experiments conducted with the goal of evaluating our proposed frameworks in chapter 3. This include education blockchain system performance evaluation, visualizations for decentralized learning data, relevance of connected learning logs and using such data to help students achieve their goals. We also discussed the implications of our solutions as well as its impact on teaching and learning.

We conclude this work in chapter 5 and report key findings from design and impact evaluations. We further discussed the implications of our work for the field, limitations and directions for future research.

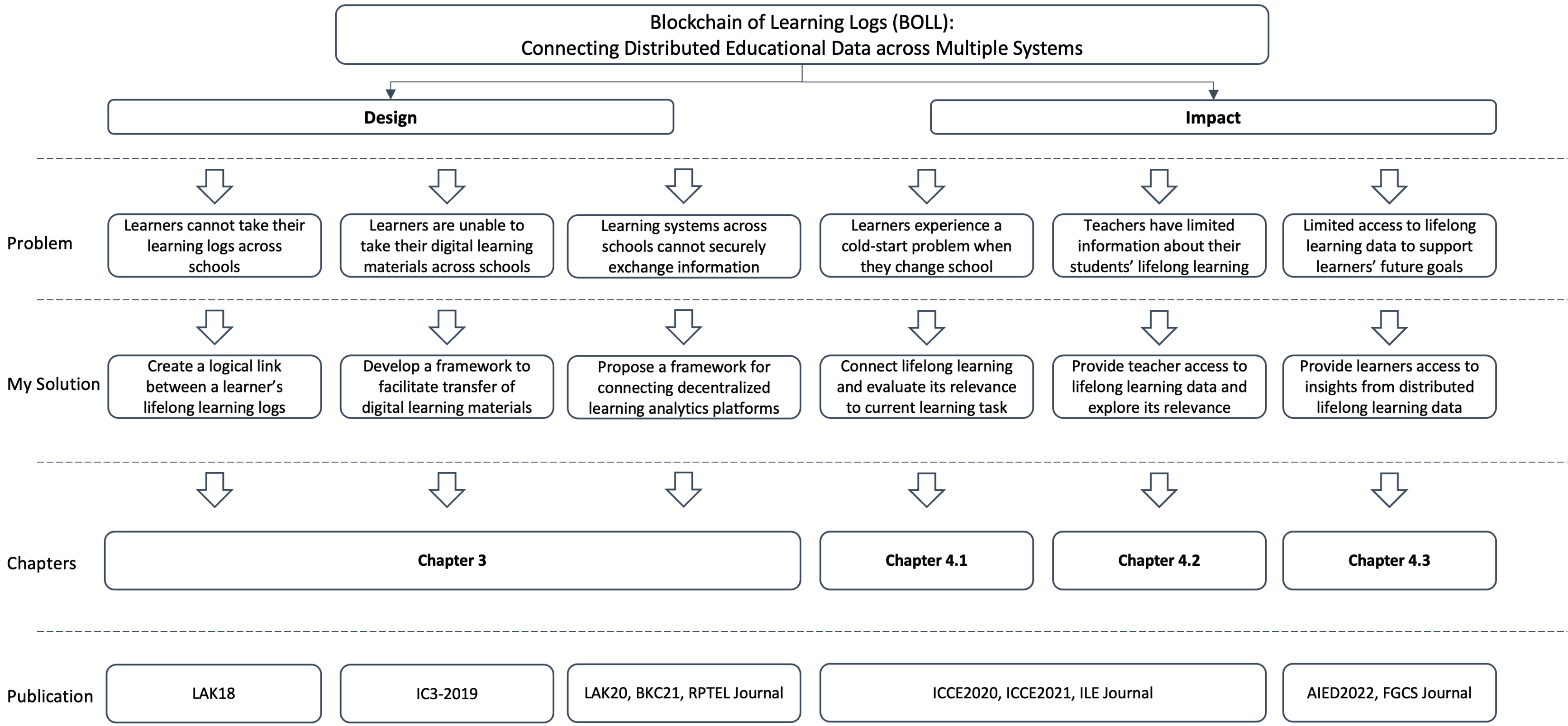


Figure 1.4: Thesis overview: Research objective, associated problems and proposed solutions

Chapter 2

Literature review

2.1 Learning technologies

There are various learning technologies that have been developed to help learners achieve their learning goals. These learning tools are found useful in teaching and learning and have had immense impact on education systems (Kanuka, 2008). In this section, we introduce some generic class of these learning technologies that are important to this work.

2.1.1 Learning management systems

Learning Management System (LMS) provide platform for instructors to deliver various learning contents to students at scale and create effective online learning communities (Beatty & Ulasewicz, 2006). In addition to the traditional classrooms, LMS's provide a means to carryout teaching activities in different ways including online learning, flipped classrooms and blended learning. A more recent way of delivering courses as Massive Open Online Courses (MOOC's) has pushed further the boundaries of delivering learning at scale. These learning technologies have contributed in making quality education available to more people breaking the previous barriers such as cost, location, gender and race (Emanuel, 2013). Fundamentally, LMS's and MOOC platforms share a common idea of delivering online learning. They both provide a way to capture events representing the learning activities of learners while using these platforms. The experiment conducted in this thesis make use of the learning logs of students on a Moodle LMS: an open source learning content management system (Moodle, 2001).

Apart from the LMS and MOOC platform, there are other learning tools for facilitating online learning. For instance, (Flanagan & Ogata, 2017a) designed a digital book reader called BookRoll. Different from regular ebook reader applications, BookRoll provide students with interactive features such as highlighting concepts as easy, needs explanation or unclear, and adding memos or taking notes. BookRoll can also render quizzes, and even provide an interface for responding to non-multiple-choice questions like essays and math problems. Another interesting feature of BookRoll and probably the most exciting for learning analytics, is the ability to record all interactions student make with any digital book viewed on BooKRoll. These logged interactions range from simple activities like page turns and time spent on a page to more complex activities such as pen strokes and delay intervals between strokes while solving a math problem. These learning actions provide a mechanism to better understand how students interact with books, lecture slides or other learning materials on BookRoll. In this thesis, learning logs of students who used BookRoll were collected and used in experimenting the proposed framework.

2.1.2 Learning data management

Data management is an important consideration in deploying technologies that envisage many users leading to more generation of data. For learning technologies that intend to capture various learning activities of learners, it is important to develop a proper method to handle and store their data. Learning Record Store (LRS) is a tool that fosters the storage of learning records including test questions, learner's responses, test scores, learning material interactions, and other learning activities. LRS's provide a means to separate the data layer of learning technologies from the presentation and service layers. A good attribute that is expected of most LRS's is the conformance of the stored data to a known standard especially if the reuse of the stored data by other learning technologies is expected.

Fortunately, LTI standards such as xAPI and IMS Caliper provide specifications for reporting learning statements. These standards can easily be adopted in the implementation of an LRS. Note that LRS is not a new type of database. In fact, an LRS stores its data in a transactional database (e.g. MySQL) or an object store (e.g. Mongo database). LRS's play a major role in ensuring that the stored data conforms to a particular format, validates the data source and then writes the data to the database in a standardized format. In this thesis, an open source LRS called the OpenLRW (Open Learning Records Warehouse) (Apereo, 2016) is used to demonstrate how data captured from an LRS can be written to the blockchain. The OpenLRW ensures that learning statements conform to the IMS Caliper standard before storing them in a Mongo database. It also provide utilities for converting xAPI statements to IMS Caliper format.

2.2 Blockchain technology

A blockchain is a decentralized and distributed peer-to-peer network which has a single immutable public ledger containing all transactions performed by participants on the network. As with decentralized systems, no single authority on the blockchain has the complete control on the activities of all other participants rather, everyone takes part in validating and processing all transactions sent over the network. While previous research works play a fundamental role in the current implementation of the blockchain, the first concrete implementation was proposed by Satoshi Nakamoto in 2008 as Bitcoin. Bitcoin is a peer-to-peer electronic cash system where online payments can be sent from one party to another without a mediating central authority (Nakamoto, 2008).

The bitcoin blockchain use cryptographic computations to ensure that the elimination of a mediating central authority does not lead to dishonest behaviors on the network. Transactions on the network are represented as connected blocks of digital signatures (see Figure 2.1) where the current transaction being considered for processing contain a hash of the previously processed transaction, the public key of the recipient, a signature of the sender and other information.

2.2.1 Types of blockchain

There are different types of blockchain: public, private and consortium blockchain as shown in Table 2.1. In a public blockchain everyone can participate in the mining of blocks and the ledger entries are public. A private blockchain restricts access within the group and the rules of the network are often determined by the convener. Whereas in a consortium blockchain, access is restricted within the group but everyone in the group has equal voting rights and decisions are made by consensus.

This thesis recommends that the proposed BOLL system should be run as a consortium blockchain. One reason for this recommendation is to allow institutions to easily claim their stake and join in determining the rules of the network. By being able to claim their stake on the network, each institution can provide as much resources to mine the blocks of their students thereby contributing to the integrity and performance of the entire network.

Table 2.1: Comparison of the different types of blockchain Zheng et al., 2017

Property	Public blockchain	Consortium blockchain	Private blockchain
Consensus determination	All miners	Selected set of nodes	One organization
Read permission	Public	Public or restricted	Public or restricted
Immutability	Almost impossible to tamper	Could be tampered	Could be tampered
Efficiency	Low	High	High
Centralized	No	Partially	Yes
Consensus process	Permissionless	Permissioned	Permissioned

2.2.2 Consensus on the blockchain

To process a transaction on the blockchain, the transaction is first signed by the sender and broadcast to all parties on the network referred to as nodes who then assert the authenticity of the transaction: the sending party must truly own the amount to be spent and such amount is not already spent in another unprocessed transaction. The latter which is known as the double spending problem hardly happens because all nodes on the network are aware of all transactions. However, if a node receives first, the transaction following an unprocessed transaction, that node will in error begin to process this false current-transaction with a wrong balance of unspent sum. Fortunately, the bitcoin blockchain resolves this by ensuring that before a transaction is added to the public ledger, all nodes must compete (a process known as mining) to solve a computational puzzle known as the Proof of Work (PoW) with all nodes agreeing on the solution once one is found. This method of reaching an agreement on a decentralized network is commonly referred to as consensus algorithm. A brief discussion on the proof of work and other consensus algorithms will be presented next.

A. Proof of work (PoW)

The Proof of Work is a cryptographic puzzle that involves finding a value whose SHA-256 hash begins with a given number of zero bits. The computational complexity of this problem is exponential to the number of leading zero bits and only nodes with faster CPUs or GPUs can solve this puzzle faster. To incentivize nodes to take part in the Proof

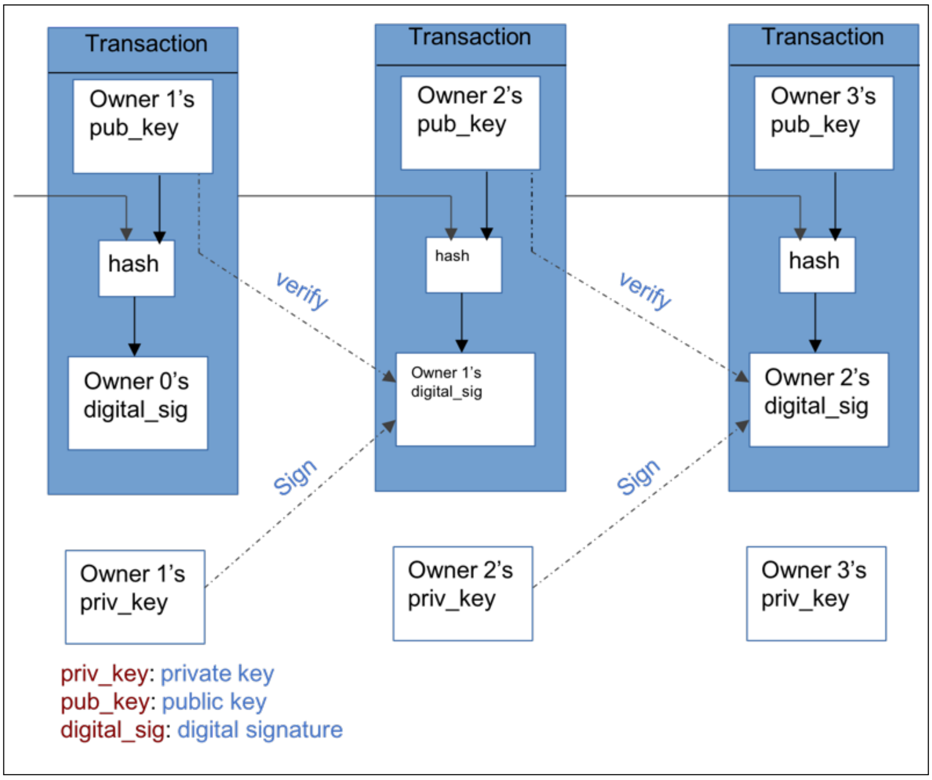


Figure 2.1: Transaction chaining on the blockchain.

of Work computation, a transaction fee known as gas is attached to every transaction. The first node to successfully compute this puzzle broadcasts the solution on the network and upon successful verification by all other nodes, the successful node is awarded the attached transaction cost and the transaction is added to the blockchain.

To ensure that only one or very few nodes compute the Proof of Work faster than others, the difficulty of the puzzle is reviewed according to the computational strength of all the nodes. In some rare cases, more than one node may solve the puzzle at the same time and broadcast to all nodes on the network. But, a node can only receive a solution from only one node and rejects the others. In this case, one would expect that the network will be out of sync and different nodes will have different ledger entries. However, because transactions on the blockchain are 'chained' as shown in Figure 2.1, potential disparity in ledger entries can be resolved upon receiving subsequent transactions where only transactions that lead to the formation of the longest chain of entries on the ledger will be favored over those of shorter length. The experiments conducted in this thesis used the PoW consensus algorithm as a way to obtain the worst computational complexity when all members take part in the mining blocks.

B. Proof of stake (PoS)

One inherent limitation with having all the members of the network compete for the right to mine the next block is the amount of time and resources required to effectively mine a block. To overcome this limitation of the PoW consensus algorithm, the Proof of Stake algorithm proposes that a miner can mine the next block if they can prove that they own as much on the network. This rides on the idea that members with more stake on the network are unlikely to attack the system. Also, to avoid a monopoly when one member becomes richer than others, a mechanism of randomly selecting the miner from a set of rich miners may be used. Because the PoS algorithm does not consider mining cost, attacks can be less difficult to orchestrate on the network. However, the gains from using PoS algorithm include energy savings, fewer validating nodes, faster mining and a potential for distributed computation of blocks.

C. Delegated proof of stake (DPoS)

This is similar to the PoS algorithm. While all stakeholders in a PoS blockchain can take part in mining the next block, DPoS algorithm allow stakeholders to nominate or delegate the mining task to another member on the network. The delegated member could be a stakeholder or an ordinary member and delegates can be removed or reassigned easily. The advantage of the DPoS consensus algorithm is that block validations would require fewer nodes and consequently, mining can occur faster.

D. Proof of authority (PoA)

Proof of Authority algorithm is a consensus algorithm that delegates the mining of blocks to elected nodes (De Angelis et al., 2018). As the name implies, the miners are known and trusted authorities which inherently eliminates the need to verify identity of miners. Thus, PoA achieves a better performance. However, one limitation of the PoA's usage in a consortium blockchain is the tendency of selected authorities with a majority vote hijacking the network to perform unauthorized operations and even voting other members out. In an educational setting where institutes are regulated by the government, it is necessary to put in place additional authorities that can neutralize the effects of such a mutiny on the network. In the proposed deployment of BOLL as a consortium blockchain (discussed in Section 2.2.1), the PoA consensus algorithm is recommended. The likelihood of attack is reduced if for instance, members of the consortium acting as stakeholders are known.

2.2.3 Decentralized applications (DApps) on the blockchain

A Smart contract is a cryptographic "box" that contains value and specific rules in a state transition functions which only unlock the value if the specified rules are met (Wood et al., 2014). These specific rules could be terms of agreement between two or more parties or rights to ownership of a commodity. The basic idea behind smart contracts is to make it possible to embed into a hardware or software different kinds of contractual clauses including collateral, bonding, delineation of property rights, etc. such that it will make the breach of contract expensive for the violator (Szabo, 1997). Thus, with smart contracts on the blockchain, one can easily specify conditions that are required to be met before a transaction is processed. The blockchain easily enforces the ground principle of making breach of contracts expensive through its transaction irreversibility feature. For an attack to be successfully carried out on the network, it is required that the attacker controls up to 51% of the entire network.

Smart contracts are the backbone of DApps. MedRec (Azaria et al., 2016) for instance, uses the smart contract feature on Ethereum blockchain to implement a decentralized system where medical records of patients are stored on the blockchain and patients can easily grant or revoke access to different health institutions. In a practical sense, the actual medical records are not stored on the blockchain due to the limit on the amount of data that can be sent with a transaction and the consequent computational implications. Rather, a reference to the location (database or API endpoint) of the actual medical record and the access authorization specifications are defined in the smart contracts.

2.3 Blockchain in education

Educational institutions and learning organizations have also found innovative ways to use the blockchain technology to control access and sharing of various assets and resources as reviewed by (Bracamonte & Okada, 2017; Chen et al., 2018; Grech & Camilleri, 2017; Sharples & Domingue, 2016). While at the time of writing there were few applications within the field, (Grech & Camilleri, 2017) suggest that there are many potential aspects of the education sector in which blockchain can be used such as: *multi-step accreditation,* *recognition and transfer of credits, rewarding use and re-use of an intellectual property, and students funding and payments on the blockchain.* Also, (Ocheja, Flanagan, Ueda, et al., 2019) reported that most education blockchains have failed to consider academic records different from certificates. Although certificates are useful for administrative and enrolment decisions, it is difficult to conduct learning analytics on such documents. Thus, to understand a learner's knowledge depth and preferences, it is important to enable connection of more granular data such as learning behaviour logs, quiz questions and scores, learning materials, and other data collected by learning tools. The blockchain is very useful in making this connection as it allows protection of the learners' data, conformance to rules defined by different institutes on the network and a decentralized management of records among multiple institutes.

2.3.1 Lifelong learning passport on the blockchain

A. Certificates on the blockchain

Schmidt (Schmidt, 2016) proposed Blockcerts, originally from Open Badge (The Mozilla Foundation & in collaboration with The MacArthur Foundation, 2012), as an open standard for creating, issuing, viewing, and verifying blockchain-based certificates. Cryptographic hashes of these certificates are stored on the blockchain where they are protected from malicious alteration and unauthorized access. However, the granularity of learning process is important for learning analytics and achieving data-driven education. As certificates are mainly a representation of accomplishments and do not express the process of learning, the proposed system in this thesis considers not only certificates, but also fine grained learning log data.

Another project by the University of Nicosia (UNIC) (University of Nicosia, 2014) looks at also placing academic certificates of its students on the blockchain. While this is similar to Blockcerts, UNIC operates from a more specific angle of single institution use case, and the certificates are for courses taken on its Massive Open Online Course (MOOC) platform. Another difference between Blockcerts and UNIC's implementation is that the former stores hashes of certificates distinctly, while the latter groups reference to certificates for students in a particular course term and store the hash of the grouped references on the blockchain. Although the approach of grouping certificates together may be advantageous for storage optimization, it might become a limitation for access and privacy where a single cryptographic hash points to multiple students' certificates. The system proposed in this thesis solves this limitation of UNIC by completely storing learning records distinctly, and allowing third party tools to determine the relationships between learning logs.

Blockcerts and UNIC's MOOC platform provide open mechanisms for verifying educational records stored on the blockchain. However, these applications are yet to provide a way to manage permissions to these records on the blockchain. On Blockcerts and UNIC's MOOC platform, permissions that define access to educational records are being managed outside the blockchain. This method of managing permissions to data is referred to as "off-blockchain authorization". One of the problems with this type of permission management is that it requires a potential accessor to look for a means (undefined means) to contact the owner of the data to request access. For example, if a researcher wants to run a survey using anonymous data on the blockchain, the researcher will be required to first know about the existence of the participant before requesting access using no specifically defined means. Consequently, the anonymized data on the blockchain becomes de-anonymized as the researcher already knowns the authorizing party. In contrast, BOLL provides an on-blockchain authorization where permissions to educational records can also be managed on the blockchain by the definition and use of smart contracts. This makes it possible for learners to remain anonymous while granting access to their learning records unless they willfully provide additional information to the accessor.

B. Microcredentials and Microbadges

Microcredentials are credentials that represent competencies, skills and learning outcomes derived from assessment-based, non-degree activities with verifiable evidence of the content of the earned achievement (Ehlers et al., 2018). Upon completion of a microcredential, the learner receives a digital badge often referred to as a microbadge as a proof of learning experiences offered by the microcredential. A microbadge contain information such as the issuing institution, requirements for earning the badge, evidence that the learner has met these requirements and the date the badge was earned. While this form of credentials and skill acquisition have emerged in recent times, they are different from the focus of this thesis. Microcredentials similar to certificates contain only summary information and proof of learning experience. They do not contain granular information such as the learner's learning behaviour, cognitive actions, preferences, or learning difficulties. Also, microcredentials do not have any form of logical links that could enable learning traceability even though they are offered digitally and collectively managed by the learner. This thesis provides a means for learning institutions to offer not just microcredentials and microbadges but also to logically connect the associated log data of each learner that can provide a data-rich environment for learning analytics and enable learning traceability.

C. Lifelong learning on the blockchain

Sony Corporation and Sony Global Education (Education, 2017) published a press statement about a system already developed to apply blockchain technology - IBM Blockchain powered by Hyperledger Fabric 1.0, to the field of education. This system is said to have two core functions; authenticate and control usage rights of educational data and an application programming interface for handling these rights aimed at educational institutions. While the goals of the ideas expressed in the press release are similar to ours, Sony Global Education is yet to publish any technical document on the implementation or usage specification of this system.

To the best of my knowledge, this thesis is the first to provide a technical specification on the application of blockchain technology to educational records different from certificates. In this work, the key contribution is to provide a framework for connecting decentralized learning logs and deliver a concrete implementation of a blockchain-based platform using the proposed framework. This thesis show that it is possible to achieve a privacy-preserving lifelong learning log using the blockchain with defined smart contracts, discuss resource requirements, and the benefits of the proposed system. A discussion on potential challenges that may be faced and possible solutions on how such issues could be tackled are provided in this thesis.

2.3.2 Learning content access and transfer on the blockchain

There are many previous works on the need to protect Intellectual Property Rights (IPR's) in a digital world including (Anderson et al., 2003; Chuang & Sirbu, 2000; Council et al., 2000; Foroughi et al., 2002; Liu et al., 2003). Foroughi et al. (Foroughi et al., 2002) discussed the use of policing, litigation, restricted sharing, and the use of Digital Rights Management (DRM) elements including content encryption, keys, passwords, and third-parties for tracking usage to ensure non-violation of digital rights. Anderson et al. (Anderson et al., 2003) proposed an eXtensible Access Control Markup Language

(XACML) geared towards achieving more usability of digital assets over a broad spectrum of applications and to also ensure security policies defined by asset owners are adhered to. To achieve this, XACML provides a request and response format for interacting with the policy system and how to interpret such policies. A Policy Decision Point (PDP) evaluates applicable policy and renders authorization decision while a Policy Enforcement Point (PEP) performs access control by making decision requests and enforcing authorization outcomes. Lorch, Proctor, Lepro, Kafura and Shah (Lorch et al., 2003), demonstrated how XACML can be used by distributed systems to achieve a more robust access control. XACML is observed to overcome the limitations of Shibboleth (Cantor & Scavo, 2005) such as Shibboleth's dependence on htaccess files which are inherently deficient and not so easy to share in arbitary locations (Lorch et al., 2003). However, XACML and the implementation in (Lorch et al., 2003) does not provide a mechanism for engendering trust between two or more potentially distrustful parties without the need for a central authority to act as a mediator.

To solve the problem of lack of trust and eliminate the need for a third-party, Zhu et al. (Zhu et al., 2018) proposed a Transaction-based Acces Control (TBAC) assets management system on blockchain which is fundamentally built on an Attribute-based Access Control (ABAC) model(V. C. Hu et al., 2013). Using the Bitcoin blockchain, Zhu et al. showed how a digital asset can be escrowed on the blockchain and protected with policies defined in state functions. While the ideas proposed by Zhu et al. are similar to ours, we find their work limited in handling multi-party scenarios such as a sponsoring organization providing access to learning resources to a learner and only pays for what the learner actually uses (parties involved: *sponsor-author-learner*). Another difference is a scenario in academic where one or more authors may write a learning resource together but each of the authors would like to manage access or changes to their contributions differently. Thus, an education-specific implementation of IPR's management becomes even more necessary as educational assets are frequently accessed, updated, and constitute different kinds of data that engender further analytics by not just the asset owner but also the accessor (learner or her institution). Also, while the goal of an asset management system might be to control access and ensure policies are not abused, we consider the case of students learning, sharing and accessing learning resources in digital forms to be peculiar and thus, should be handled differently.

Also, we consider implementations of e-learning systems and/or marketplaces such as these (Abelson, 2008; Emanuel, 2013; Thompson, 2011) to be limited in facilitating interactions between potentially distrustful parties and the lack of transferability of learning footprints across different institutions. Hoffman et al. (Hoffman et al., 2018) and Janowicz et al. (Janowicz et al., 2018) also identified the possibility of using a decentralized network to offer IPR's protection in education but focused on using the blockchain to manage journal management workflows.

We overcome these limitations of enabling access control and tracking intellectual rights violations by providing a distributed network of authors, publishers and users where intellectual works can be shared over the network with a publicly held history of all access. Using smart contracts, we make it possible for authors or publishers to define clauses that should be met before access is granted. With this design, no third-party is required and users can ensure their privacy is not breached by choosing what should be tracked or not and acceptable to the content provider. It should be noted that this is different from the traditional understanding of agreements. While offline agreements may require one time sign-off, the smart contracts on the blockchain are constantly queried to ensure that both parties agree to the terms already defined within the contract with no party having the autonomy to change its content without informing the other party.

The Blockchain of Learning Logs (BOLL) proposed in (Ocheja, Flanagan, Ueda, et al., 2019) enables the realization of lifelong learning logs for students as they move from one learning environment to another. Using the blockchain technology, BOLL ensures that students' privacy is protected through permissions defined in installed smart contracts. The BOLL framework forms a fundamental background for our work. Our proposed framework for transferability of learning materials allows broad auditing by concerned parties on the network and also permits digital content owners to decide how their contents from the Data Depository Server (DDS) are served to other users in order to facilitate better policy violation tracking. Also, to improve learning outcomes, we introduce a mechanism for users to rate and recommend useful contents to one another.

2.3.3 Education blockchain data visualization and analytics

Specific reviews on education blockchain data visualization and analytics research is still lacking. One could argue that this is so because the use of blockchain in the field of education is still new and even the concept of learning analytics suffers low adoption (Macfadyen et al., 2014; Tsai & Gasevic, 2017). We consider our contributions vital to setting the field on the right path as various learning organizations begin to adopt the decentralized paradigm to managing academic records. This work also provides further clarity to how education blockchain systems can be situated in core learning processes so as to impact directly teaching and learning outcomes. Through the data on the blockchain, we present visualizations where teachers can access lifelong learning records of their students held at different schools, measure students' readiness for learning new concepts, discover missing learning blocks and actively support their students in a more personalized manner. For students, we present visualizations that will enable reflection on learning experiences across different contexts and learning environments, provide interfaces for revising and reviewing past learned concepts, and at the same time place students in control of their privacy and data protection. Administrators are also presented with visualizations where they can view how academic records on the blockchain affect learning, data protection policies adherence or infringements, and awareness of students' engagement with learning resources provided by the institution.

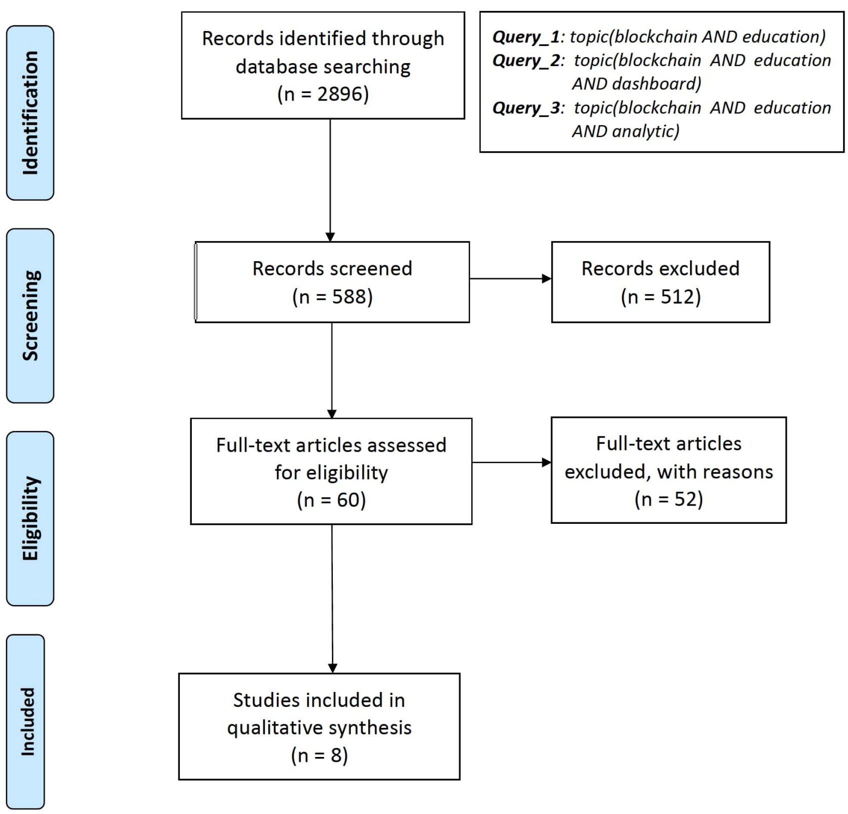


Figure 2.2: Data collection process PRISMA diagram ((Moher et al., 2009))

Data collection and results

To understand the current trends in visualization of education data on the blockchain, we set out to review research related to visualizations in education blockchain. We searched multiple academic databases including Web of Science, IEEE Xplore Digital Library, ACM Digital Library, MDPI database, and other sources (Google Scholar, Researchgate, Archive, etc.) for publications within a 10-year period (2010-2020) and written in English language. We search these databases using a combination of queries and following the prisma guideline for systematic reviews (Moher et al., 2009). As outlined in figure 2.2, after initial evaluation from the title of the records, abstracts review and analysis of the scope of some of these papers, only 60 papers were considered for full reading. Results from *Query_1* were excluded as it was too broad: returning over 2,000 results. For the search results from *Query_2* and *Query_3*, we only considered papers on education blockchain that presented proposals or working implementations that included visualizations or user interfaces for the various features of the system. Only 8 works met these criteria and were included for review.

Classification scheme and methodology

Dashboards are the most common way to communicate information on learning systems. LAD's as previously reviewed can be used to create awareness, enable reflection and engender behavioural change towards improved performance or taking actions to achieve specific goals. Information such as scores, grades, statistical distribution of various cohorts, engagement overview and behavioural analysis can be viewed on most LAD's. In reviewing how education blockchain systems contribute to these objectives of LAD's,

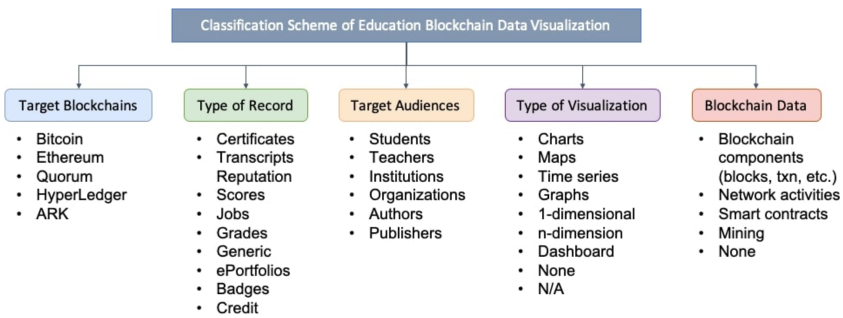


Figure 2.3: Classification scheme of education blockchain data visualization.

Application	Target Blockchain	Type of Record	Target Audience	Type of Visualization	Blockchain Data
EduCTX (Turkanović et al., 2018)	ARK, Ethereum	Credit (ECTS)	Students, institutions, private sector	1-dimensional	None
Blockcerts (Schmidt, 2016)	Bitcoin, Ethereum, HyperLedger	Certificates	Students, institutions, Organizations	1-dimensional, dashboard	None
EduRSS (H. Li & Han, 2019)	Ethereum	N/A	Students and institutions	1-dimensional, graphs	Blockchain components
QualiChain (*)	Ethereum	Certificates, badges, ePortfolios, jobs, reputations, tutor reservation	Students, teachers, institutions, organizations	Dashboard, maps	None

* Kontzinos et al., 2020; Mikroyannidis, 2020; Mikroyannidis et al., 2018

Figure 2.4: Classification of education blockchain data visualization.

in figure 2.3, we categorize the various implementations into five (5) categories: target blockchains used, type of academic records managed on the blockchain, target audience, the different types of visualizations, and the blockchain data that can be found on the visualizations. We note that most of the reviewed works did not provide information consistent with LAD's rather they present user interfaces with data about the user's information or system features. These user interfaces do not provide engaging features for learners or teachers to support their learning or teaching objectives as seen in other LAD's. A summary of the results from our review is shown in figure 2.4. What follows is a case-by-case analysis of the education data visualization features provided by each implementation of blockchain in education. Also, the visualizations from these systems are shown in figures 2.5, 2.6, 2.7 and 2.8.

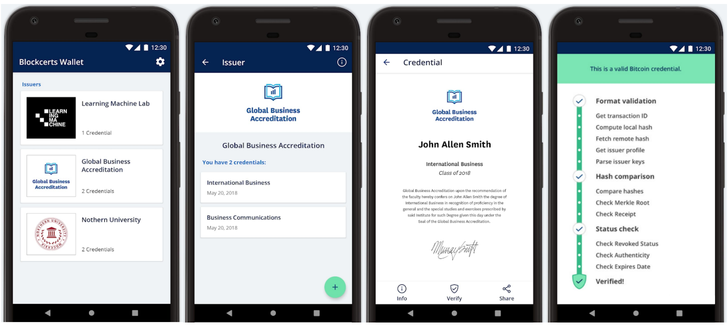


Figure 2.5: Blockcerts wallet showing list of issuers, and certificate detail (Schmidt, 2016).

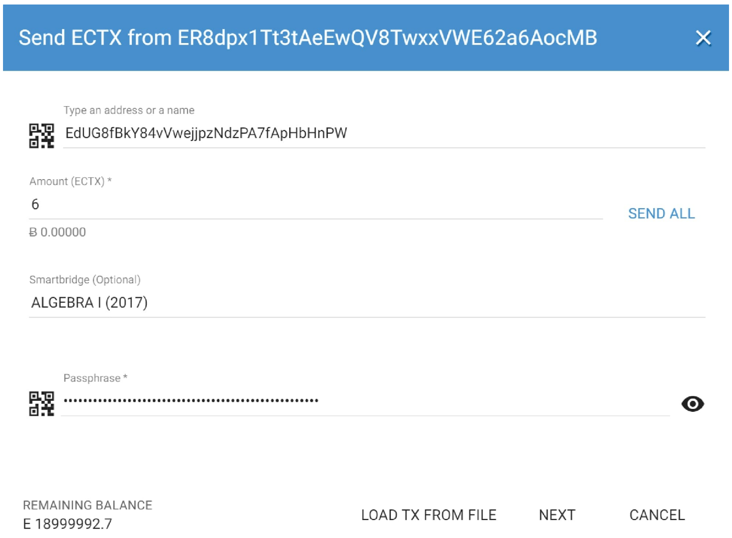


Figure 2.6: Professor assigning credits using the ECTX client wallet.

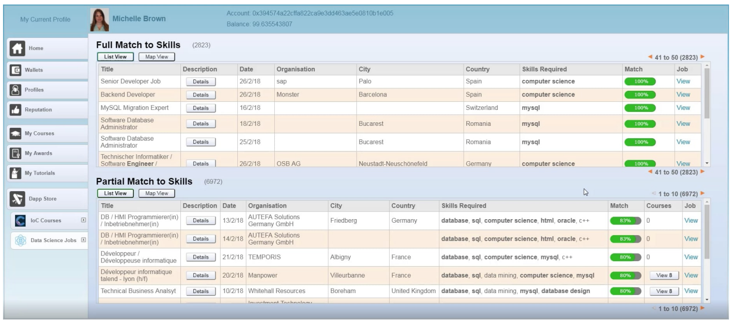


Figure 2.7: Jobs listing on QualiChain (KMi, 2020; Mikroyannidis, 2020).

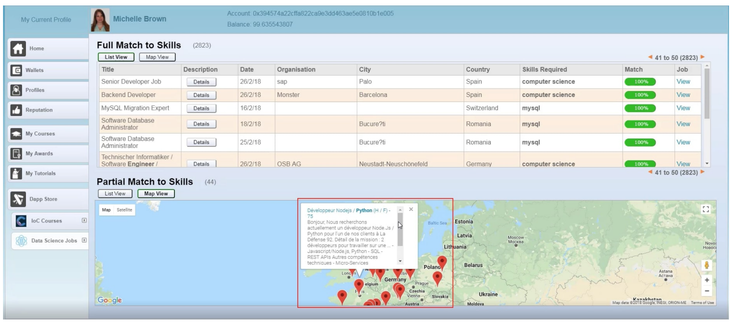


Figure 2.8: Jobs listing on QualiChain (map view) (KMi, 2020; Mikroyannidis, 2020).

2.4 Distributed learning analytics

2.4.1 Standards and interoperability of learning tools

One visible limitation with too many learning technologies is the problem of interoperability: how can learning records from one learning tool be correctly interpreted by another learning tool? To solve this problem, different learning organizations have proposed standards to foster Learning Tools Interoperability (LTI). For instance, the Apereo Initiative proposed the Experience API (xAPI) standard for reporting or making a statement about a learner's activity within a learning tool. Using a structure similar to the Resource Description Framework (RDF) ontology (RDF triple: *actor*, *verb* and *object*), the xAPI standard specify how learning statements should be reported such that the semantic and contextual meaning of learning records are correctly conveyed. Figure 2.9 show a sample xAPI statement. The actor refers to the learner or user who carried out this learning activity. The *verb* attribute describes the action that was performed by actor while the object refers to the entity that received the action. With this understanding, we can read the sample xAPI statement in Figure 2.9 as follow: *admin launched a book.*

Another standard similar to the xAPI is the IMS Caliper standard (Consortium, 2013). While xAPI focuses on reporting any type of learning experience or evidence in various forms (online or offline), the IMS Caliper standard focuses on providing quantitative metrics for learning, real-time data messaging, and give details on student engagement in learning activities. However, both standards follow similar RDF-triple ontology and a bidirectional conversion is possible. In a similar manner, the framework proposed in this thesis modeled equivalent smart contracts to accommodate known LTI standards that follow the RDF-triple ontology like the xAPI and IMS Caliper. The smart contract for learning records relies on action-defining aspects of the learning records for proper grouping of similar learning records.

2.4.2 Lifelong learning analytics

In this section, we evaluate the state-of-the-art on enabling lifelong learning for learning analytics. Due to limited solutions that can facilitate learning data continuity across different learning environments, we examine previous attempts at enabling and using data from multiple sources to improve learning. We model the act of combining data from multiple sources as a common alternative when lifelong learning data of learners

```
{
    "version" : "1.0.0",
    "actor" : {
        "objectType" : "Agent",
        "account" : {
            "name" : "admin",
            "homePage" : "https://www.example.org/bookroll"
        }
    },
    "verb" : {
        "id" : "https://w3id.org/xapi/adb/verbs/launched"
    },
    "object" : {
        "objectType" : "Activity",
        "id" : "https://www.example.org/bookroll/book/view?contents=123",
        "definition" : {
            "name" : {
                "en" : "Moodleとは"
            },
            "description" : {
                "en" : "a1b2c3"
            },
            "extensions" : {
            ...
            }
        }
    }
}
```

Figure 2.9: An example xAPI statement.

is difficult to obtain or store. While there are previous reviews on multisource data for learning analytics as reviewed in (Samuelsen et al., 2019), this section is focused on lifelong learning and analytics: identifying common traits of existing solutions, enabling learning traceability, connecting learning experiences, revisit consistent stakeholder concerns on data privacy and security and present vital pivots for steering the field towards enabling connected lifelong learning data.

In this review, we adopt Kitchenham and Charters (Kitchenham & Charters, 2007) review guidelines consisting of planning conducting and data synthesis. Planning this review is informed by the need to evaluate the progress of enabling lifelong learning. While there are no previous works that specifically review advancements in enabling lifelong learning, A review of other works such as (Ferguson, 2012; Flanagan & Ogata, 2017b; Samuelsen et al., 2019; Siemens et al., 2011) showed that lifelong learning is desirable and useful for learning analytics. This argument is further strengthened by the recurrent use of data from different sources (Samuelsen et al., 2019) such as social media to perform learning analytics. Thus, this review sets out to answer the following questions:

1. What is the state-of-the-art infrastructure lifelong learning?
2. How has the field evolved in using lifelong learning or multisource data for learning analytics?
3. Privacy and Stakeholders: what are the threats to adoption of new technologies for enabling lifelong learning?

To conduct the review, we also focus on searching the ACM Digital Library containing the proceedings for Learning Analytics and Knowledge (LAK) conferences from 2011 to 2019. To get a good coverage of various trends in the use of multisource data or lifelong learning for learning analytics, we also include publications and systematic reviews on lifelong learning and multisource data for learning analytics available on other databases such as (Misiejuk & Wasson, 2017; Samuelsen et al., 2019). In table 2.2, we show the criteria for including a paper published in LAK 2011 - 2019 according to the PRiSMA guideline (Moher et al., 2009) shown in Figure 2.10. For example, research works that use multisource data or a distributed system are considered for review.

Table 2.2: Inclusion Criteria

Criteria	***Description***
Distributed	The study use multisource data or the system used in the study must run in a distributed manner.
Transfer	The paper accesses or transfers learner data across multiple institutions or parties for learning analytics.
Interaction	The research consider interaction between students across different institutions or learning environments.
Ethics	Studies that discuss ethical issues with multisource data or transfer of learner data.
Implementation	If the study is a framework or platform, a concrete implementation should be available.

On the other hand, some studies which might have some of the inclusion criteria are excluded according to the exclusion criteria in table 2.3.

A total of 84 works were identified from the databases scanned. 28 papers were screened out based on the tables 2.2 and 2.3. Currently, 8 full-text articles have been

Table 2.3: Exclusion Criteria

Criteria	***Description***
Collaboration	The study focuses on collaboration between actors in the same learning environment (not distributed).
Generic	The work has a more specific focus which varies significantly from the inclusion criteria.

accessed and found eligible for inclusion in qualitative synthesis. We do not focus on in-depth quantitative synthesis as metrics such as count of implementations or design methods do not provide credence to solving the problems with lifelong learning.

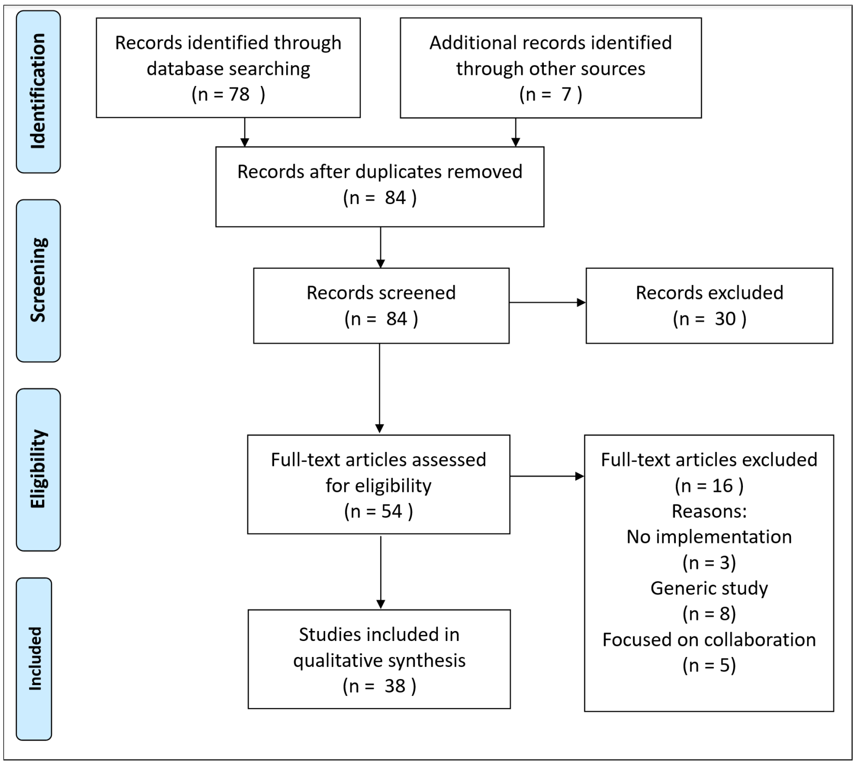


Figure 2.10: Review process diagram.

2.4.3 Results

We perform a thematic analysis (Braun & Clarke, 2006) on the selected papers under themes: Infrastructure (Systems, Analysis and/or Specifications for implementation), and Governance and Ethics (e.g. stakeholder perception and adoption, and privacy).

A. Infrastructure

Papers on Infrastructure addressed challenges of fragmented learner data, variation in data formats which inhibits interoperability of learning tools, data granularity, learning environment peculiarity, adoption, performance, scalability, data collection, selection, analysis, visualization and distribution.

For example, at the first LAK, Suthers and Rosen (2011) proposed a method to abstract and unify learning analytics processes using the analytic hierarchy containing event models with representations. The reason for modeling such an abstract transcript is to facilitate its usability across different environments. For example, the process trace event model looks at learner's interaction process on learning environments such as log data, multimedia contents and textual transcripts. The limitation with this method is that it proposes one tool for everyone. In contrast, BOLL processes everyone to their own tools but let us share the data or our results.

Kump et al. (2012) proposed how to use MyExperience to view what a learner knows as constructed by the Knowledge Indicating Event knowledge map. Details of data or data source used for the KIE were not provided.

Bienkowski et al. (2012) proposed Learning Registry as an infrastructure that supports learning resource discovery, sharing and amplification. Learning Registry attempts to provide answers to the questions on effectiveness of learning resources, target audience and what learning resources result in better performance. This is realized by storing metadata that describe learning resources. Some important aspects for consideration include user profiling, experience, and domain modeling and personalization. However, one limitation to this approach is that it is required that learning resource owners have to converge at a central point to list their learning resources. Also, this work only provide access to locations of learning resources: it does not store or facilitate sharing of learning interaction data with these learning resources. This is because these learning resources and the learner interaction data remain on the premises of the providers or owners.

While Okada and Tada (2014) uses a real world, context-aware mechanism to provide a personalized learning experience for learners in a collaborative task, it does not bother on how learning actions are stored in a sharable or reusable manner. We also agree that this work is more related to evaluating learners' performance in a learner personalized collaborative task. But we found it important to review as it shows an example of a scenario where lifelong learning data is desirable but difficult to obtain. Okada and Tada (2014) used various technology tools to capture real-world learning including wearable devices. This task would have been less difficult and reusable if lifelong learning data were captured, stored and shareable in a distributed manner. Similarly, Bakharia et al. (2016), Jiménez-Gómez et al. (2015), Kennedy et al. (2015), Mouri and Ogata (2015), Rienties and Toetenel (2016), and Van Inwegen et al. (2015) show the need for lifelong learning data for learning analytics.

Piety et al. (2014) Performed a high-level survey on four communities working with educational data: academic/institutional analytics, learning analytics/educational data mining, learning analytics/personalization and system/instructional improvement. Piety et al. (2014) envisage a time where these fields will converge into a broad field of educational data sciences where educational data can be re-usable for multiple purposes with consequent evolution of technologies, innovations, policies and impacts on education. Again, we find this work useful as it speaks to the emergence of an interdisciplinary access and use of educational data. These four communities will require data fluidity, consider ethical concerns and actively engage stakeholders in order to advance the education as a whole.

Kitto et al. (2015) presented a toolkit for data extraction from social media and data import to an LRS that supports xAPI standard. This work also states the importance of using learner data from other environments for learning analytics. However, one limitation with Kitto et al. (2015) is the need to perform such data import on case-by-case basis and an expectation of consistency of the social media platform on which scraping is done. Kitto et al. (2015) does not also address the need to connect learner data in other institutions. Mandran et al. (2015) proposed DOP8: a data mining iterative cycle which takes into account data' life cycle and operators' life cycle. While DOP8 closes the gap between both cycles, a proposed implementation called UnderTracks, facilitates managing multi-source data using a central database which is partly agnostic to data format (mandatory fields exist). The limitation of such centralization could range from privacy concerns to issues with administrative burden of maintaining and regulating such a centralized infrastructure. However, we note that the aspect of creating a link between data life cycle and operator life cycle is laudable.

Hickey et al. (2015) was organized to discuss potentials of Open Badge The Mozilla Foundation and in collaboration with The MacArthur Foundation (2012) and facilitate its adoption for representing learning accomplishments. However, a limitation that still exist with Open Badges is its focus on top level or summary statements rather than access to actual learning statements. Also, the Open Badges is a standard which in itself does not contain specifications for connecting other lifelong learning: few implementations like Blockcerts Schmidt (2016) focused on reporting and verifying certificates. Similarly, Bakharia et al. (2016) discusses the complexity with using the xAPI standard while (Whyte et al., 2016) focused on IMS Caliper Analytics specification. Bakharia et al. (2016), Hickey et al. (2015), and Whyte et al. (2016) provided useful suggestions for advancing the interoperability of learning accomplishments and statements.

Mangaroska et al. (2019) proposed a framework that enables cross-platform analytics to enhance personalization and learning adaptation. While this is similar to Ocheja et al. (2018), Mangaroska et al. (2019) does not provide details on how to collect data from different platforms. Ocheja et al. (2018) and Ocheja, Flanagan, Ueda, et al. (2019) proposed and implemented a decentralized architecture where LRS's containing learner data are connected together through the blockchain and privacy permissions are managed using smart contracts. In this work, we present how Ocheja et al. (2018) and Ocheja, Flanagan, Ueda, et al. (2019) can serve as a recipe for enabling lifelong learning without displacing existing learning tools. We also present the implications for the field over the next decade.

B. Governance and ethics

At various learning organizations, the task of deciding what learning tools to adopt are often administratively decided. While the administrative setup varies from one institution to another, the goals of learning and improving student performance are often similar. We review papers that address issues concerning methods or problems that have hindered adoption of learning technologies and analytics. The reviewed papers also provide possible solutions that may aide faster and more effective adoption of learning analytics. We use the ideas from these papers to provide meaningful ideas that could guide the field over the next decade.

Fournier et al. (2011) argue that it is useful to conduct learning analytics on online networks as it will provide stakeholders with values such as insight on meaningful interactions and actions in learning environment that could be used to improve the learning processes and outcomes. Also Fournier et al. (2011) recommend the use of both qualitative and quantitative analysis to achieve better understanding of the learning scope and get detailed information while performing analytics on online networks.

Graf et al. (2012) highlighted some specific questions relating to data governance in the education sector such as data ownership, required permissions, audit trails, how privacy constraints scope of learning analytics, and how should learning analytics be considered or regulated as a research. These questions are important considerations especially when we have enormous amount of data resulting from learning actions, complexity of relationships among institutions and the different diversity in tools used for learning and analytics.

Also Drachsler and Greller (2012) through a set of questionnaires revealed that most stakeholders agreed that learners and teachers are the main beneficiaries of learning analytics resulting also in a higher impact on teacher - student relationship. In this light, we argue to further improve teacher - student relationship, it will be useful to convey a student's past performance by the previous teacher to their current teacher through connected lifelong learning. However, Drachsler and Greller (2012) reported the low awareness of the survey respondents of the concept of learning analytics. This means there is a need to constantly engage and educate stakeholders about learning analytics.

The first Learning Analytics Readiness Instruments (LARI) survey (Arnold et al., 2014), the governance and infrastructure component bothered on the need to access an institute's readiness in the aspect of technical infrastructure, policies, institutional governance and oversight. Another factor, data, measured the extent to which different types of data are collected at the institution and how such data is managed. While Arnold et al. (2014) noted the small sample size as one of its limitations, factors of governance, and infrastructure and data were rated higher than the average. In a further study where LARI was deployed, Oster et al. (2016) agreed that it is important to understand all characteristics associated with the implementation of learning analytics including issues on policies, abilities and infrastructure. It is therefore necessary to continuously engage researchers and stakeholders in various discussions on how to advance learning through learning analytics.

Tsai and Gasevic (2017) reviewed eight policies for learning analytics as enforced by two main groups: support organisations and research consortiums, and higher education institutions.The results which included privacy protection and data management and governance, Tsai and Gasevic considered it necessary to deal with issues on data anonymity, informed consent and options to opt out of data collection. On data management and governance, Tsai and Gasevic pointed out the need for transparency, and a clear awareness of how the collected data will be used. It is therefore recommended again that institutional leadership should be actively involved in the conversations on deployment, use and management of learning analytics at institutions. While institutional leadership may be appropriate for deciding implementation of learning analytics, it is also important to understand the expectations of students and concerns (Prinsloo & Slade, 2015; Slade et al., 2019; Whitelock-Wainwright et al., 2017).

A key aspect in learning analytics is the control of personal and private information by an individual. This includes the ability to opt out of learning activity tracking and giving parents of underaged learners the right to manage their dependents' learning records (Pardo & Siemens, 2014; Rubel & Jones, 2016). Usage or access to a learner's learning records should be sought from the learner and/or their institution depending on the terms of agreement between both parties or according to other defined policies. This agreement should contain clauses such as: usage policies, access authorization, storage policies, etc.

To solve the problem of low adoption of learning analytics at scale, Dawson et al. (2018) proposed a complexity theory-based leadership model to facilitate adoption of learning analytics. In their proposal, Dawson et al. examined existing leadership structures in the Australian higher education context to find evidence of complexity leadership. This revealed the existence of two classes of learning analytics leadership where Class 1 used an instrumental approach of adoption: learning analytics was viewed as an instrument to solve identified challenges. Class 2 viewed learning analytics as a process for informing the continuous improvement of student learning practice in response to an identified problem. However, the survey concluded that while most adoption of learning analytics start in small scale (course or class levels), the organization level of adoption of learning analytics is important. We argue that to facilitate such organizational level of adoption, it is necessary to assure senior administrators that the processes are transparent and can be conveniently managed. Using a decentralized architecture, it is possible to facilitate both small scale (course level) and large scale (organizational level) adoptions.

In a more recent effort to address policies for facilitating adoption of learning analytics, (Tsai et al., 2018) proposed SHEILA (Supporting Higher Education to Integrate Learning Analytics) policy framework. SHEILA was proposed to address learning analytics adoption challenges including: demand on resources (data, infrastructure, human and financial resources), ethics and privacy. In the results of the survey conducted with 3 groups using the six ROMA (RAPID Outcome Mapping Approach) model, dimension 2 thrives on the understanding that learning analytics implementation requires collective effort while dimension 4 corroborated the need for considering resources, ethics and privacy, and stakeholder engagement. This work shares in the goals of the SHEILA policy framework of actively engaging all stakeholders in the process of facilitating lifelong learning. Specifically, the use of a decentralized network for lifelong learning ensures that no stakeholder is being overlooked and the resources requirement for advancing lifelong learning can be distributed over the network. With respect to the financial burdens of managing a decentralized network, we argue that institutions who have financial constraints can have the option of partnership with other institutions or government agencies to realize these objectives at lesser cost.

2.5 Summary of gap in previous work

In this chapter, we have summarized previous work and their limitations. From a design perspective of connecting lifelong learning and analytics, the following problems were identified:

1. Learners are unable to connect their learning logs across different schools.
2. Learners are unable to transfer their learning materials when they change school.
3. It is difficult to analyze multisource data and supporting technology for distributed learning analytics is still lacking.

From the evaluation perspective, there were limited research on the use of blockchain to connect education data apart from certificates and transcripts. Also, existing proposals on connecting academic data on the blockchain do not have concrete implementations and lack evaluations on usefulness to teaching and learning, usability, feasibility, resource requirements, and other affordances.

In addition to the literature review, we also conducted exploratory study and needs analysis to identify stakeholders' perception on the relevance of accessing past learning records of learners. Further details and results from the exploratory study are presented in Chapter 4.2.

Chapter 3

BOLL framework: connected lifelong learning and analytics

3.1 System architecture

In this thesis, we propose the Blockchain of Learning Logs (BOLL) that can connect academic records of learners across schools, enable transfer of digital contents and provide visualizations and tools to support lifelong learning and analytics. In Figure 3.1 we show an overview of the Boll system and the unique features that solve problems earlier identified with enabling connected lifelong learning and analytics including security, traceability, transferability, verification, decentralized access, and immutability. What follows is a brief description of each component (pie segment) of Figure 3.1.

The security component of Boll facilitates decentralized authentication across different learning tools, protects user privacy through anonymization and access control by data owners. The traceability component ensures that lifelong learning remains traceable, accessible and can be analyzed for additional insights to support various learning goals. Transferability module allows students to move with their data across different schools including digital learning materials and lecture slides and continuously manage permissions related with such transfers. The verification component provides a means for third party data consumers to check ownership and the validity of logs reported on the Boll network. The final components: data storage and sources provide interfaces to connect existing learning tools and data silos to the Boll network.

In the following sections, we discuss our research efforts in designing and implementing the Boll system. These works have been peer-reviewed and published in international journals and conferences.

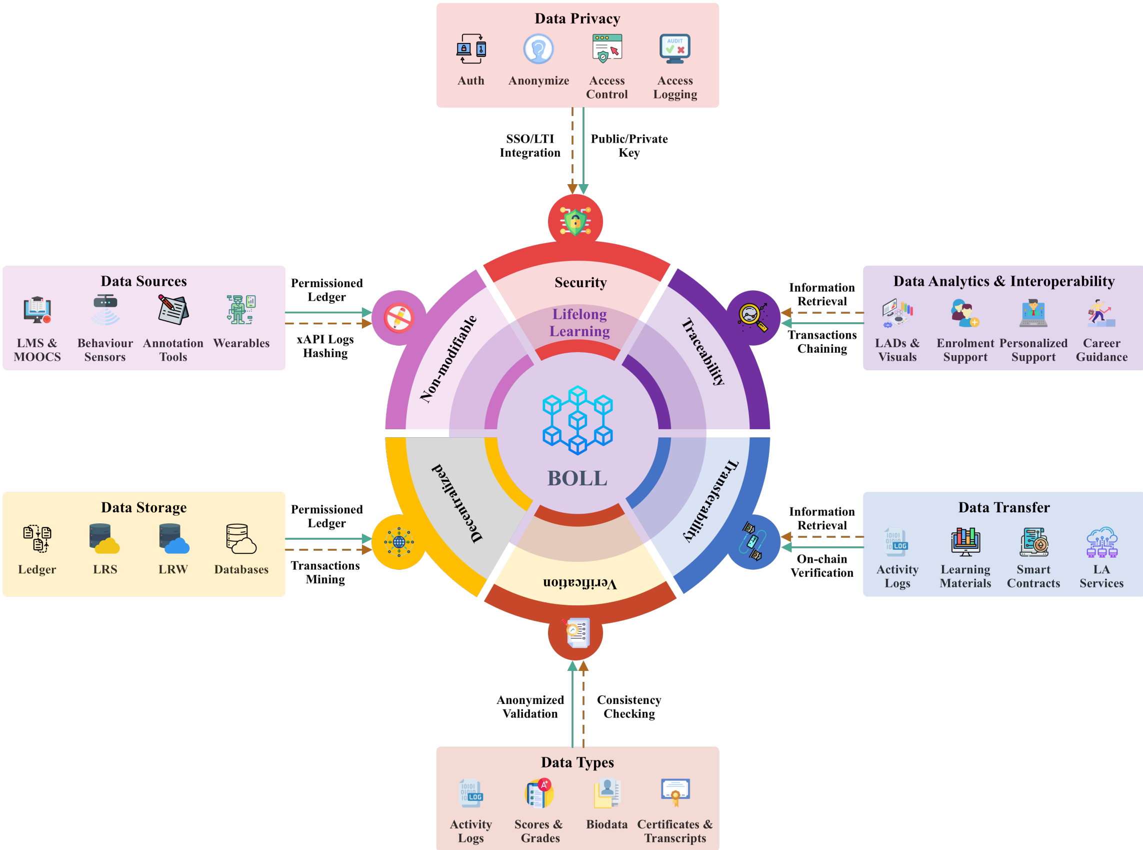


Figure 3.1: Boll system architecture.

3.2 Connecting distributed learning logs

3.2.1 Overview

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 (Learning, 2016a), IMS Caliper Discovery API (Consortium, 2015) 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.

As institutes maintain separate LRS's which are not connected to one another, this results in the learning data that was collected at previous institutes not being available for analysis at current or future institutes. The situation causes a typical *cold-start problem*, where the current institution's learning environment does not have sufficient data for effective personalization or adaptation when the learner is first enrolled. In this paper, we propose a solution that enables the logical movement of learning records using a blockchain as a transport medium and platform for connecting LRS's. In particular, the following problems are addressed by the proposed solution:

1. Distributed learning logs across multiple institutions caused by the use of independent disconnected LRS's.

2. Inability to transfer or access a learner's data and testimonials across multiple institutions, making it difficult to achieve lifelong learning logging.

3. The lack of protection and control of private information by data owners.

Solutions to these problems is integral to the further development of the learning analytics, learning personalization and learning enhancement. A main motivation of this research is to develop lifelong learning logs for learners. A lifelong learning log typically contain verifiable proves of all the learning activities carried out by a learner (Ogata et al., 2011). As learning is a continuous and an ongoing process, Ogata et al. proposed a lifelong learning log as a personal and private journal for documenting learning activities. In this work, we present an implementation where such a journal is recorded as a secure entry on the blockchain. The authenticity of the journal can also be verified easily by consensus using the data stored as blocks within the blockchain, and could be used in assessing a persons educational achievement, suitability for employment, and intellectual evaluation. We are particularly interested in using the blockchain to solve this problem because it provides a mechanism for:

- Distributed consensus, data consistency and immutability of processed transactions. These features can make it nearly impossible to alter learning records on the network (Nakamoto, 2008).

- Defining clauses or contracts on the blockchain that determine how learning record data transactions are handled and protected.
- Facilitating interaction between multiple stakeholders (institutions, students, 3rd parties) with high transparency and protection of each participant's interest as agreed in defined contracts.

It is also important to protect private information while enabling lifelong learning logs. A key aspect in learning analytics is the control of personal and private information by an individual. This includes the ability to opt out of learning activity tracking and giving parents of underaged learners the right to manage their dependents' learning records (36, 2016; Pardo & Siemens, 2014; Rubel & Jones, 2016). Usage or access to a learner's learning records should be sought from the learner and/or their institution depending on the terms of agreement between both parties or according to other defined policies. This agreement should contain clauses such as: usage policies, access authorization, storage policies, etc. Our proposed solution is to facilitate these agreements on the blockchain by allowing the learner and their institutions to act as signatories on defined smart contracts, and enable the protection of learning records on the blockchain.

Although different institutions utilize different learning platforms, some standards have been proposed for enabling learning records from one institution's learning platform to be correctly interpreted on another institution's platform (Consortium, 2013; Learning, 2016b). These standards along with those proposed by Ocheja et al. (2018) were used in this paper to implement a system for connecting learning records generated at different institutions on a single public ledger.

3.2.2 Architecture design

We propose a Blockchain of Learning Logs (BOLL): a blockchain platform that connects the learning logs of students across the different institutions they have attended on a single, public and immutable ledger. We present BOLL as a solution to the problem of transferring educational data between different institutions as students move from one institution to another. It also solves the cold-start problem in learning analytics systems where a new students' learning environment is created without being informed by previous learning activities, even though their current learning activity is based on experiences at their previous school. Previous learning data could serve as a robust foundation upon which new learning environments are created when a learner enrolls in a new institute. Ocheja et al. (2018) identified key features of the blockchain that makes this implementation possible. These include decentralization, single public ledger, privacy, immutability and the deployment of smart contracts. We build on these key features to enable connected learning logs across different institutions, defined smart contracts to regulate access, and implement mechanisms to classify learning logs to also enable easy indexing and quick lookup times.

Currently, various institutions and learning platforms store and manage their learning records separately with no standard method to move learning records from one platform to another without duplicating user information as shown in Figure 3.2. In Figure 3.3, we propose a change from current implementations of learning management systems and platforms to a blockchain of learning logs where all learning institutions can co-exist on a single public ledger. This can be facilitated by using the proposed BOLL system and policies defined by smart contracts. Institutions that take part in the BOLL system can agree to allow students access to their learning records while at other institutions, and can state the conditions for such access on the blockchain. The proposed system also solves the problem of different user accounts at multiple institutions by linking a single BOLL user identity to all LRS's within the network.

We also use the nested transactions feature of the blockchain where the contents of blocks represent pointers to learning data with ownership and access policies. Nodes on the peer-to-peer network represent learning providers. Learning activities performed by learners on the learning platforms of learning providers on the network are logged on the blockchain as string representations of queries that can be executed on the LRS's of

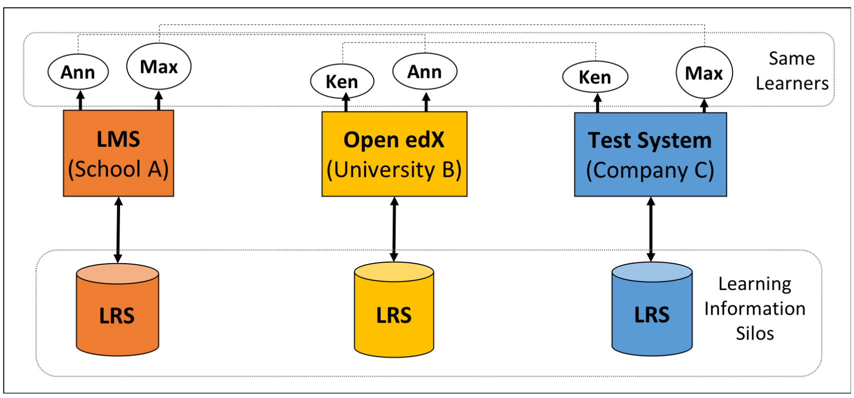


Figure 3.2: Current learning systems design.

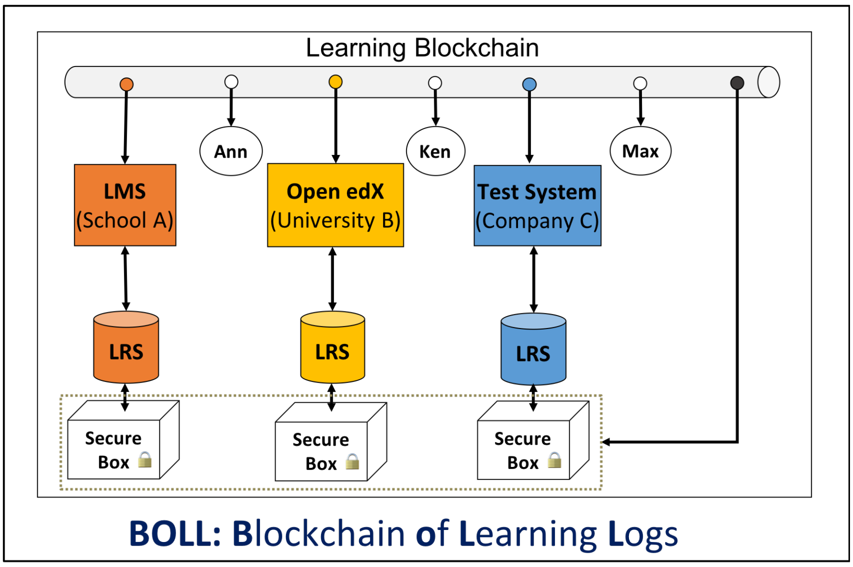


Figure 3.3: Proposed design of blockchain of learning logs (BOLL).

learning providers to retrieve such activities. To ensure data consistency and immutability, at transaction initiation time, we execute accompanying queries on the LRS and include a cryptographic hash of the 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 a secure box for executing these queries against a providers' LRS's with reference to the blockchain network to maintain established permissions.

Figure 3.4 shows a typical setup of our implementation for one institution. We use an open-source Learning Management System (LMS), Moodle (Moodle, 2001) and a digital book reader, BookRoll (Flanagan & Ogata, 2017a) as the learning tools. All learning records emitted from these tools through learning activities of learners are stored in a central database, MongoDB: a document-oriented database, through Open Learning Record Warehouse (OpenLRW) which is an open-source LRS (Apereo, 2016). These learning records are either in conformance with the xAPI standard (Learning, 2016b) or IMS Caliper standard (Consortium, 2013). We also provide an implementation of a subroutine for retrieving records from the MongoDB through a wrapper method on OpenLRW and writing them to the blockchain. For this implementation, we used the open source Ethereum blockchain written in Go programming language (Ethereum, 2013a).

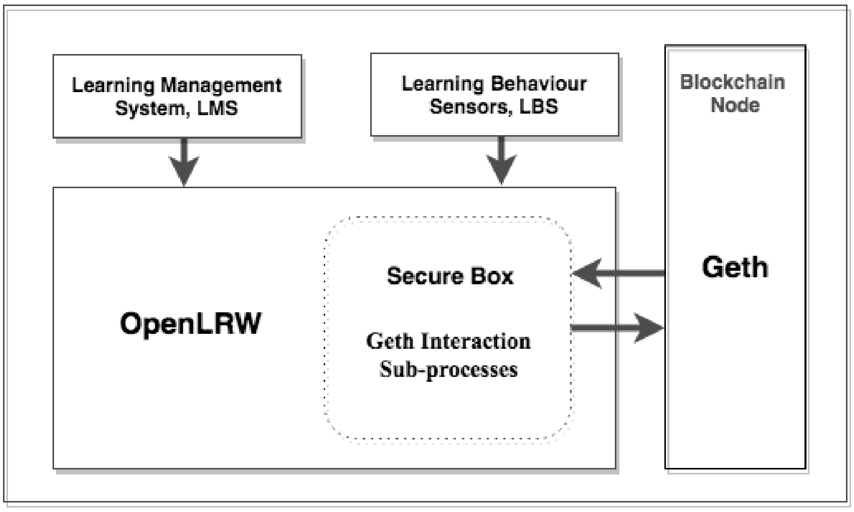


Figure 3.4: BOLL Architecture - one institution.

3.2.3 Smart contracts schema

The BOLL system enforces smart contracts that contain learning data access permissions, ownership, and a mapping between the permissions and ownership. The state transition functions of these smart contracts can be modified to reflect the conditions that should be met before data read or write access is granted. Figure 3.5 shows a hypothetical hierarchical design of these smart contracts. We define three main smart contracts namely: Registrar – Learning Provider Contract (RLPC), Learner – Learning Provider Contract (LLPC) and Index Contract (IC) for both Providers and Learners.

The RLPC controls how students, teachers, organizations and institutions become authorized learners or learning providers on the learning blockchain. For institutions and organizations, these requirements are often administratively decided. Hence, we propose that typical implementations should consider existing structures for establishing communication and accessing information in institutions and organizations. In our implementation, we maintain a registry of institutions that are allowed to join the BOLL system's network using the institution's domain name and an encrypted message signed with their private key and then verified with their public key. In Table 3.1, we describe some of the attributes/functions defined in the RLPC.

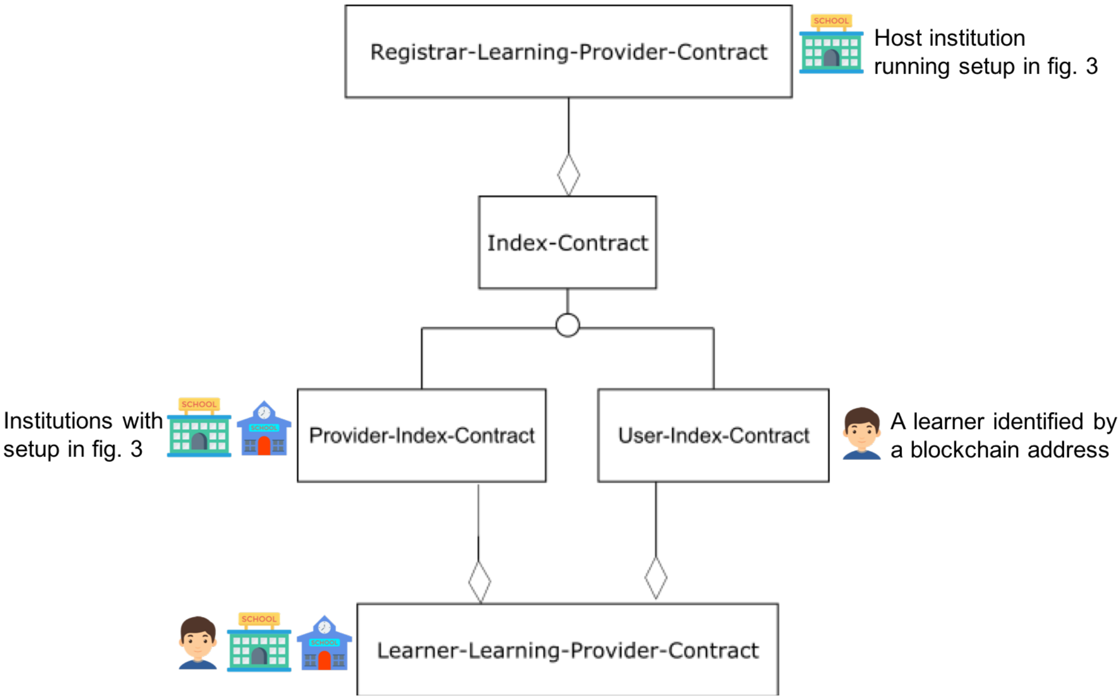


Figure 3.5: Hierarchical view of BOLL system smart contracts.

Table 3.1: Registrar-Learning Provider Contract - RLPC

Attribute	**Description**
owner	address of starting institution
registered participants	addresses of participants mapped to their Index Contract
register	a function for registering new users
unregister	deactivates a user
assign Index Contract	assigns an Index Contract to a user

An Index Contract is also installed to provide a mechanism for fast look-up of entries and access permissions on BOLL. This unique design solves the current limitation of Solidity(Ethereum, 2013b), which is a smart contract programming language that does not provide a look-up interface for data types. Although, arrays and hash table-like data types are provided, the cost of looking-up an entry in an array computationally grows with the size of the array. On the other hand, the hash table-like implementation does not provide an interface for accessing the keys to the values in the hash table. This means that, to look-up any entry in the hash table, we should have the key stored elsewhere.

In our case, to look-up a learning event, we should have a pointer to that learning event. Thus, it is necessary to develop a mechanism for storing the pointers or keys to the learning events, otherwise we risk losing information written on the blockchain. In Table 3.2 and 3.3, we define the internal contents of the Provider Index Contract (PIC) and the User Index Contract (UIC) respectively. The PIC is for learning providers while the UIC is peculiar to learners. We use a hash table-like implementation for keeping a list that maps learners to their LLPC's, and another list that maps learning providers to LLPC's they have with learners and with other learning providers that learners have granted access.

The LLPC represents a proof of existence of a learner's learning data on a learning provider's platform. This smart contract is dedicated specifically to handling a learner's learning record and how it is accessed. We decided to use a specific smart contract for this purpose so as to make it easy to transfer learning records from one institution to another. With our design, a transfer can easily be done by invoking the *grantAccess* function (with permission from the owner or their institution) on the LLPC without erasing or physically dislodging the learning record. The LLPC contains: information such as the blockchain address of the owner, the URL of the originating learning provider's LRS with a hashed id parameter for retrieving the original record, a hash of expected learning data for ensuring data has not been tampered with and a key-value pair of institution's address and their access permissions (read, write, grant-read, grant-write, none).

Table 3.2: Provider Index Contract - PIC

Attribute	**Description**
owner	address of institution
learners to learning records	a mapping of learners' address to their LLPC's
learners	a list of all learners at this institution
insert learning records	inserts a new LLPC
get learning records	retrieves a learner's learning records

Table 3.3: User Index Contract - UIC

Attribute	Description
owner	address of learner
providers to learning records	a mapping of providers' address to learner's LLPC's
providers	a list of a learner's learning providers
insert learning records	inserts a new LLPC
get learning records by providers	retrieves LLPC written by a learning provider
get learning records by record type	retrieves LLPC for a given action verb

The address refers to a hexadecimal string uniquely generated and having corresponding private and public keys. A learner can have as many LLPC's as the number of distinct types of learning events carried out. These events could be any of the xAPI or IMS Caliper action verbs (Consortium, 2017; Learning, 2016b). Also, an institution may request access to read a student's learning logs contained in an LLPC contract by invoking the *requestAccess* function in the LLPC smart contract. Other invocable functions on the LLPC smart contract as shown in Table 3.4 include *insertLearningEvent, grantAccess and revokeAccess* which respectively insert learning event, and grant or revoke access to a learning record.

Table 3.4: Learner-Learning-Provider Contract - LLPC

Attribute	Description
owner	address of learner
record type	the action verb for this series of learning events
permissions	mapping of providers to their allowed permissions; *read, write, grant*
learning events	list of learning events of the same record type
insert learning event	adds a new learning event
request/grant/revoke access	*ask/give/deny* access to this LLPC

table continued on next page

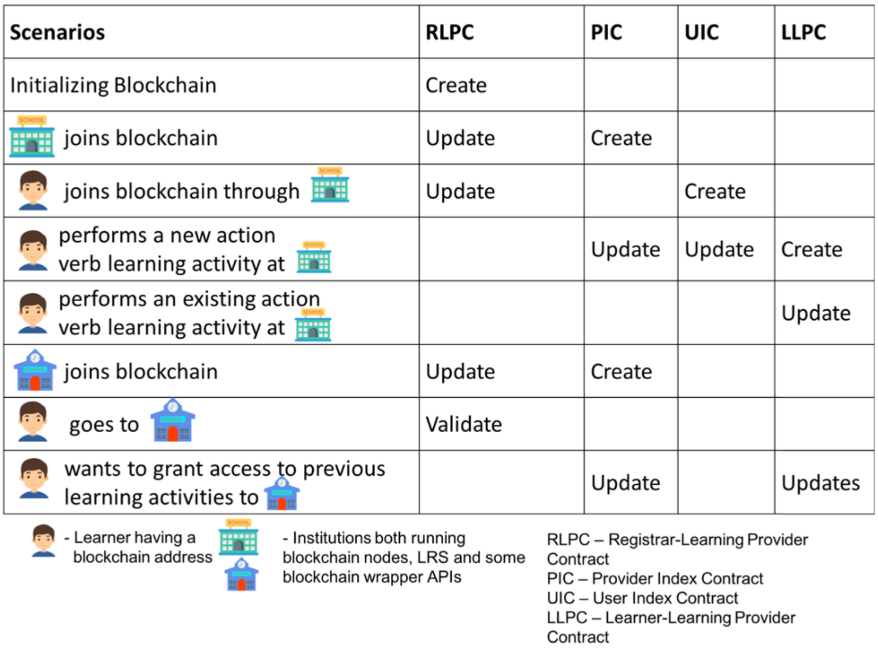
table continued from previous page

Attribute	Description
pending requests	a collection of pending access requests

3.2.4 User interface design

A BOLL system setup consists of at least one institution as shown in Figure 3.4 to serve as host. BOLL has two main user groups: institutions, and teachers/learners/students. We will now discuss the required steps in setting up a BOLL. In Figure 3.6, we make a list of processes, actors and the necessary smart contracts with the required operations to be performed. The RLPC is first installed on the blockchain node serving as the host. One institution should volunteer to serve as the host node. With this, all institutions that wish to join the blockchain will have to request to be registered by having a similar setup as in Figure 3.4, and then sending a registration request to the RLPC which was initially installed on the hosting institution's blockchain node. Upon approval, the RLPC is updated with their information and a PIC is created. 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 – UIC and the final phase of registering the generated blockchain address and UIC address in the RLPC.

On the blockchain, learning records are uniquely grouped using the action verb field and the user's blockchain address. Writing learning histories involves 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 as we want to optimize gas usage on the blockchain. Gas as used here refers to the computational cost for processing transactions on the blockchain. The amount of gas required to process a transaction increases with the size of the data in the transaction to be processed. Hence, 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 a 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 need to be mined on the blockchain.



Scenarios	RLPC	PIC	UIC	LLPC
Initializing Blockchain	Create			
joins blockchain	Update	Create		
joins blockchain through	Update		Create	
performs a new action verb learning activity at		Update	Update	Create
performs an existing action verb learning activity at				Update
joins blockchain	Update	Create		
goes to	Validate			
wants to grant access to previous learning activities to		Update		Updates

- Learner having a blockchain address
- Institutions both running blockchain nodes, LRS and some blockchain wrapper APIs

RLPC – Registrar-Learning Provider Contract
PIC – Provider Index Contract
UIC – User Index Contract
LLPC – Learner-Learning Provider Contract

Figure 3.6: Processes involved in enrolling or accessing information on BOLL.

3.3 Connecting distributed learning materials

3.3.1 Overview

As the amount of data in the digital space continue to grow leading to more meaningful use cases, it is important to ensure appropriate use, reward ingenuity and foster collaboration among diverse parties. Intellectual Property Rights (IPR's) are rights that allow creators or owners of industrial properties (patents for inventions, trademarks, etc.) or copyrighted works (books, poems, artistic works, etc.) to benefit from their own work or investment in a creation by defining terms of usage which potential users of their work should comply with (May, 2006). With many works on the use of technology to solve issues relating to IPR's protection and enforcement such as (Anderson et al., 2003; Cantor & Scavo, 2005; Hoffman et al., 2018; Janowicz et al., 2018; Lorch et al., 2003; Zhu et al., 2018), we focus on specific issues on how educators including students and teachers can share learning materials in a secure, privacy-enabled, intellectual rights-aware and collaborative environment.

In various e-learning environments, it is common to have teachers share learning resources such as slides, lecture notes, books, quizzes and assignments with their students. Students could in turn also make meaningful use of these resources to arrive at new resources that other students or the teacher might find helpful. With more knowledge resulting from simple interactions like this, we consider it necessary to have a system that supports exchange of these resources, reward ingenuity, increase distribution, foster collaboration, and protect the intellectual rights of the authors. Thus, this paper is inspired by the need to solve these problems:

1. How do we ensure trust and transparency between an author of a work that is made available to students and a sponsoring organization that pays for the author's work based on the usage quota of each student without using any third-party?

2. How can students generate and share learning materials with their peers across different schools with IPR's protection?

3. For companies, other learning organizations, and publishers, how can we establish a trusted and transparent network where these actors can co-exist and provide a wide pool of educational resources to students?

We provide solutions to the above problems by extending the framework for a blockchain-based learning analytics platform proposed in (Ocheja et al., 2018) and implemented in (**ocheja2019**) as a Blockchain of Learning Logs (BOLL). BOLL is a decentralized platform that enables logical movement of students and their academic records from one institution to another. Different from certificates or transcripts issuing systems, BOLL provides a mechanism to share learning logs of students on the various learning tools they interacted with while studying at different institutions. Our main contributions are:

1. We engender trust between sponsors, authors, and users of their work by providing transparent auditing of access to learning materials on a decentralized network.

2. We propose algorithms for programming smart contracts that enforce privacy and IPR's.

3. We design and discuss a framework for realizing a decentralized e-learning marketplace for a healthy co-existence among parties with varied interests.

3.3.2 Architecture design

Figure 3.7 shows our proposed framework for enabling a decentralized e-learning marketplace for managing authorship and tracking access to digital contents on BOLL. BOLL Marketplace (BOLL-M) comprises of two groups of stakeholders; authors and users. Authors refer to actors on BOLL-M who own intellectual rights to learning materials made available in the marketplace. While users refer to members of the BOLL network who wish to access learning materials made available in the market and/or organizations that provide sponsorship for students to access learning materials (e.g. a government education ministry or other funding organizations). A student or teacher on BOLL-M can also be an author of a learning material in the marketplace. In this scenario, the student or teacher can rely on the learning material publishing tool made available to them by their institution. For publishers who do not belong to an academic institution, it is required for them to be authorized by the BOLL Consortium proposed in **ocheja2019**. After such an authorization is acquired, the publisher can setup a node on the BOLL network as show in figure 3.8. We will now describe each of the components shown in figure 3.7.

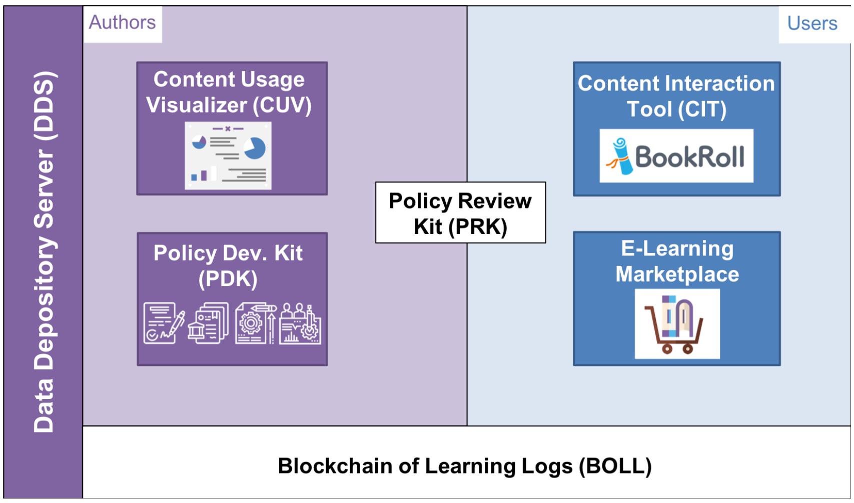


Figure 3.7: Decentralized e-learning marketplace

3.3.3 Smart contracts schema

To enable appropriate use of learning materials on BOLL-M, it is necessary to define policies that accessors should comply with. We represent these policies as state transition functions in the smart contracts. Due to the technical skills required to write smart contracts, we provide multiple templates as a Policy Development Kit (PDK) which authors

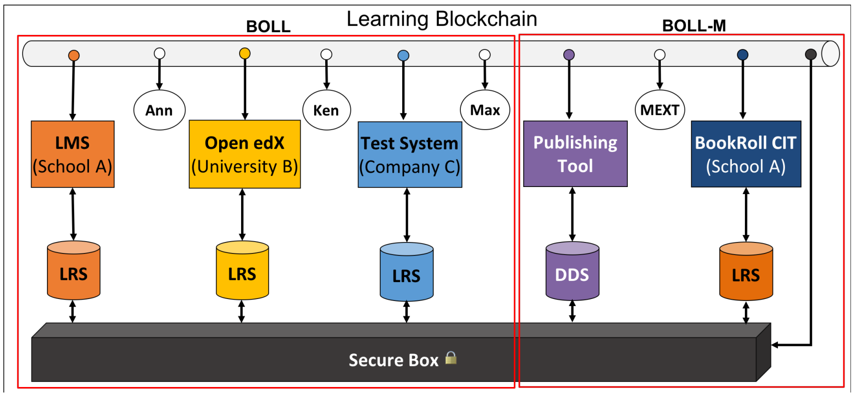


Figure 3.8: Decentralized e-learning marketplace on BOLL

can choose from, adapt to their use case and install on BOLL to protect access to their learning materials. We represent these smart contracts in four broad categories.

A. One-time signatory policy (OSP)

This refers to smart contracts that can be installed once and contain clauses on how a learning material can be accessed and used with the permission of the author. When an OSP is issued on BOLL-M, it is irrevocable and the issuer either grants a limited or lifetime access to a learning material depending on the duration specified. An example of a useful application is where students are given a one-time limited access to a professional or degree examination provided by another organization. In algorithm 1, we show a pseudo-code for issuing an OSP by an author identified by public key, Pk_{author} to a learner with public key, $Pk_{learner}$. The implementations of the $getSigner(message)$ and $notify(message)$ are not shown in this work as one could easily use the public key resolution and event emitting features of the blockchain as well.

Algorithm 1 Procedure for issuing a One-time Signatory Policy (OSP)

Pk_{author} : author's public key, $Pk_{learner}$: learner's public key Pk_{owner} : public key of the policy recipient
$getSigner(message)$: returns public key of the signer $notify(message)$: emits an event or sends a broadcast

procedure issueOSP $(Pk_{author}, Pk_{learner}, message)$
$getSigner(message) = Pk_{author}$ $Pk_{owner} = Pk_{author}$ $Pk_{owner} \longleftarrow Pk_{learner}$
notify$(Pk_{learner})$
end procedure;

Dual party signatory policy (DPSP): This is a revocable version of OSP where two parties can agree or disagree on the terms of access to a learning material. In a DPSP, terms of access can be modified by the issuer and such modified version becomes valid only when the accessor of the learning material agrees to the new terms. DPSP is useful in scenarios where an author maintains a continuously improved version of a learning material (e.g. lecture slides being updated regularly) and does not wish to create an entirely new version with the new changes. Although, smart contracts once installed are immutable, we achieve versioning of terms of access by allowing the execution of sets of instructions within the contract until all parties, S_{Pk} append their signatures, S'_{Pk}. The learner will be notified when this happens and only after that can the learner access such learning material. Algorithm 2 gives an illustration of a typical DPSP.

Algorithm 2 Procedure for issuing a Dual Party Signatory Policy (DPSP)

S_{Pk} : public keys of stakeholders S'_{Pk}: public keys of stakeholders who've approved DPSP

$access_{grant}$: indicates if DPSP is approved

procedure issueDPSP $(Pk_{author},\ Pk_{learner},\ message)$
$access_{grant} = false\ \ Pk_{author} \in S_{Pk}\ and\ getSigner(message) = Pk_{author}\ \ Pk_{author} \notin S'_{Pk}$
$S'_{Pk}[Pk_{author}] \longleftarrow 1$

$length(S_{Pk}) = length(S'_{Pk})\ \ Pk_{owner} \longleftarrow Pk_{learner}$
$access_{grant} \longleftarrow true$
notify$(Pk_{learner})$
end procedure;

B. Multi party signatory policy (MPSP)

The MPSP is a collaboration enabled smart contract that allows multiple parties to determine the conditions for accessing a learning material. To enable multi party arbitration, MPSP starts off with the proposed clauses of the originating party. Another party can review these proposals and either refuse or accept them by invoking the state transitions functions contained in the initial MPSP. The originating party is tasked with initializing the MPSP with some settings including the participating parties ($Pk_{voter_1}...Pk_{voter_n} \in V_{Pk}$), the wining strategy ($wining\_ratio$) and the tie breaker ($Pk_{arbiter}$) as shown in Algorithm 3. For example if a simple majority wining strategy is specified ($wining\_ratio > 50\%$), the smart contract becomes valid if a simple majority agrees with the stated terms. In a case where a tie occurs, the parties can propose one party ($Pk_{arbiter}$) whom they think should be the final arbiter. This party is then given the ability to override all votes and either accept or deny the approved installation of the MPSP terms. For instance, we find the MPSP useful in a three-party scenario where one party owns and provides the learning material (e.g. publisher), the second party pays for the learning material (e.g. government) and the third party is the consumer of the learning material at no cost (e.g. students). This solves the particular problem where an organization sponsors access to a learning material on behalf of the students. The tie breaker is useful in a case where the sponsoring organization is unable to ascertain the usefulness of a learning material to the student. In this case, both the sponsoring organization and the author can delegate the student to adjudge whether they find such learning material useful or not. Algorithm 4 provides a typical demonstration of the voting procedure on an MPSP.

Algorithm 3 Procedure for initializing a Multi Party Signatory Policy (MPSP) V_{Pk} : public keys of stakeholders who can vote $Pk_{arbiter}$: public key of arbiter (tie-breaker) $arbitrate_{start}$: indicates if arbitration should/has started $Poll_{open}$: indicates if voting is still open $Votes_{Pk}$: votes cast $wining\_ratio$: the minimum fraction of total votes required for victory

procedure initializeMPSP $(V_{Pk},\ Pk_{arbiter}, wining\_ratio)$
$V_{Pk} \longleftarrow V_{Pk}$
$arbiter \longleftarrow Pk_{arbiter}$
$arbitrate_{start} \longleftarrow false$
$Poll_{open} \longleftarrow true$
$Votes_{Pk} \longleftarrow \{\}$
$wining\_ratio \longleftarrow wining\_ratio$
end procedure;

Algorithm 4 Procedure for issuing a MPSP
Pk_{voter} : voter's public key $message$: a message signed by the voter $vote$: 1 (for) or -1 (against) $sgn(value)$: signum function

procedure issue MPSP $(Pk_{voter},\ message,\ vote)$
$Poll_{open}\quad Pk_{voter} \in V_{Pk}\ and\ getSigner(message) = Pk_{voter}\quad Votes_{Pk}[Pk_{voter}] \longleftarrow vote$
totalVotes $\longleftarrow \sum_{i=0}^{length(V_{Pk})} \begin{cases} & \text{if } V_{Pki} \in Votes_{Pk}. \\ 0 & \text{otherwise.} \end{cases}$
$\frac{length(Votes_{Pk})}{length(V_{Pk})} \geq wining\_ratio$
no_tie = $\geq$ wining_ratio
no_tie $approve \longleftarrow sgn(totalVotes)\quad arbitrate_{start} \longleftarrow true$
$notify(arbiter)$
$Poll_{open} = false$
notify(V_{Pk})
$arbitrate_{start}\ getSigner(message) = arbiter\ approve \longleftarrow vote$
$arbitrate_{start} \longleftarrow false$
notify(V_{Pk})
end procedure;

C. Discovery policy (DP)

In order for an author or a publisher's learning material to show in the e-learning marketplace on BOLL-M, the author is required to install a DP smart contract. This contract contains a basic information about the learning material such as title, date published, version, description, applicable smart contracts (at least one of OSP, DPSP, MPSP). Because the DP smart contract does not contain the actual learning material or pointers to it, it is publicly available to anyone on the network to access but not modify.

3.3.4 User interface design

A. Policy review kit (PRK)

The PRK contains a set of useful tools for reviewing proposed as well as installed policies or smart contracts. This include policy modifying tools like acceptance, refusal or arbitration, and learning material rating tools. The policy modifying tools are provided to ensure that other parties understand the defined terms before accepting them. Learning material rating tools are useful for helping students find contents that might be appropriate for different scenarios based on the perception of their peers or teachers.

B. Content usage visualizer (CUV)

We propose an interface for authors and sponsors to visualize the interactions users have made with their learning materials. Since all transactions on the blockchain are written to a public ledger whose contents are immutable, we realize the CUV by querying this public ledger. However, because some functions in the installed smart contracts do not modify state and thus do not lead to transactions, we consider it a necessity that all request to view a learning material should invoke at least, a payable transaction so that access histories can also be written to the ledger. This can be achieved by mandating that all functions used to check access authorizations before responding with the learning material should write on the ledger a message signed by the accessor.

C. Data depository server (DDS)

We recommend that authors or publishers should store their learning materials on a DDS. For students and teachers who might not be able to setup the publishing tool shown in figure 3.8 (Consisting of CUV, PDK, and a part of PRK), we envisage that their schools would setup a shared publishing tool and a DDS. The DDS is connected to the SecureBox proposed in Ocheja et al., 2018 and all requests sent to the DDS are verified with BOLL through the SecureBox.

D. E-learning marketplace

The e-learning marketplace is an interface that lists all learning materials published on BOLL-M. For an author's learning material to be displayed in the marketplace, it is required that the author should install a DP smart contract. This contract can be retrieved from the PDK and adapted to the author's use case. An author may also specify that their learning material can be discovered in the marketplace by only selected users.

Content interaction tool (CIT): To ensure that intellectual rights of authors are not violated, we recommend that the tool for viewing escrowed learning materials, here referred to as Content Interaction Tool (CIT), should be connected to BOLL. In figure 3.8, we use BookRoll, a digital book reader as our CIT. BookRoll traditionally logs user interactions with digital books including bookmarking, highlights, page turns, etc. We consider these interactions enough to know when a user accesses an escrowed learning material. For recording a simple interaction on BOLL-M, one can simply log an access event when the content is being served for the first time. In a case where monitoring more interactions is desired, we can listen to specific events of the CIT. As BookRoll stores user logs on a Learning Record Store (LRS), it is possible to listen to page turn events and subsequently notify BOLL-M of these interactions. We note that monitoring of the user's interactions can be an invasion of privacy. Hence, we recommend that this should only be done according to the terms of the smart contracts.

3.4 Education blockchain data visualization

3.4.1 Overview

Big data has revolutionized many areas of business ranging from search companies to e-commerce, where insights from data have driven personalization, targeted advertising, improved services and overall business growth. However, similar success has not yet been achieved in the field of education technology, and the use of data-driven education in the field is still lagging (Siemens & Long, 2011). One of the key challenges in this area is the lack of data-continuity. When students change from one institution to another, their learning data remains largely immobile, such as in the usual progression through elementary, junior-high, and high school. As institutes maintain different data stores which are not connected to one another, this results in the learning data that was collected at previous institutes not being available for analysis at current or future institutes. The situation causes a typical *cold-start problem*, where the current institution's learning environment does not have sufficient data for effective personalization or adaptation when the learner is first enrolled (Barnes & Stamper, 2008). The advent of new technologies such as decentralized and distributed systems provide new opportunities for advancing various fields towards a more data connected world. Grech and Camilleri, 2017 extensively discussed the blockchain technology, its advantages over traditional systems and also presented various use cases of blockchain in education such as academic credentials issuing and verification, credits reporting, e-portfolio, and intellectual property tracking. The blockchain has also been found useful in solving the cold-start problem by allowing access to a learner's lifelong learning data when they change school (Ocheja et al., 2018).

When data from multiple systems are being connected, one of the key concerns is how to present such data to the stakeholders without creating additional problems such as information overload, intractability, low or no meaningful use and inability to determine what data is available or what data could be used for what purpose. In this paper, we present results from review of existing visualizations in both blockchain and non-blockchain educational technologies and propose how education blockchain tools can provide better visualizations to aide learning goals. We also report the results from an initial validation of our proposal with teachers through a qualitative method.

This paper is organized as follows: In the second section, we identify and discuss related work by presenting past reviews on visualization of blockchain data, blockchain in education, visualization of education data (non-blockchain enabled) and education blockchain data visualization. The third section describes how we identify useful papers to include in our classification of education blockchain data visualization. The fourth section dwells on our classification scheme and what information may be considered important for stakeholders in the field of education. Section five contains a needs analysis of current stakeholders with respect to accessing and using distributed learning records. In the sixth section, we propose elements and designs for education blockchain data visualization of students' lifelong learning and achievements that could meet the needs identified. Finally, in section seven, we discuss some key findings, how our proposal address the problems earlier identified, open challenges and how to solve them.

3.4.2 Architecture and user interface design

In this section, we present the first of its kind implementation of a blockchain-based education data visualization platform that supports transfer and access to lifelong learning logs. Through the use of learning analytics on education records, and distributed access to these records, we demonstrate how different stakeholders in education can manage, and make sense of their past academic records or that of their students to support different goals. It is important to note that we are not advocating for a reinvention of LAD's for education blockchain data. Our main concern is that before now, education blockchain data has not reflected in most LAD's and the use of these data in advancing students' learning or teaching through self-regulated or personalized methods is still lagging. In fact, a common denominator across all systems that attempt to use education blockchain data is basically on access to academic credentials such as certificates, portfolios, and scores. We therefore advocate that LAD's should provide stakeholders with more actionable insights on academic records managed on the blockchain.

The features of our proposed framework for the visualization decentralized education data include:

- *Connect and analyze distributed learning records:* To connect learning records, we use the transaction chaining ability of the blockchain. Each time a record is generated on the learning platform, a learning record transaction is formed and written to the blockchain as per the specification in (Ocheja et al., 2018). Analysis of these records is then performed through learning analytics methods such as knowledge modelling.

- *Privacy of learner data:* Our proposed system enable learners to control who can access their learning records even after they graduate. This is implemented using smart contracts on the blockchain to manage permissions to their records. We present a visualization where learners can manage (add, edit and delete) these permissions without having prior knowledge of the blockchain.

- *System and data security:* Our proposal uses the consensus algorithm of the blockchain to engender trust between different parties. Also, our implementation added some custom events which are emitted each time specific transactions are processed for a user. Users can use the provided interface to indicate whether they should receive notifications for such transactions or not.

- *Tracing learner learning path:* Given a single learning record belonging to a learner, our proposed framework can validate its correctness and retrieve all preceding, and succeeding records. This is achieved using the inherent merkle tree structure of transactions on the blockchain. One way of visualizing this learner learning path trace is discussed later in the subsection on decentralization and data sharing.

An important question that we must first answer is this: What does blockchain bring to education that makes it so peculiar in comparison with current learning systems and platform? Education blockchain mainly brings the features of the blockchain to the education space. These features include: decentralization, sharing of data across multiple systems, traceability, privacy, and security. For each of these features, we will discuss how they affect education systems and visualizations on LAD's. In Section 4.2.4 to 4.2.7, we propose and implement visualizations that can improve data awareness of various stakeholders taking into account the above features.

3.5 Decentralized learning analytics

3.5.1 Overview

Evaluating students and reporting performance outcome is an essential component in teaching and learning. However, information on students' performance are often not readily available to decision makers (teachers, students, parents, etc.) or even provided in a comprehensible format (Zapata-Rivera & Katz, 2014). This problem makes it difficult for students to use their performance data to measure how they compare to their peers at other institutions especially when enrolling into higher education. The root causes of these problems range from limitations such as data privacy, interoperability, lack of distributed analytics and consequent implications of such distributed access (Baker et al., 2019). Thus, we propose a framework to enable a decentralized reporting and access to assessment results based on a Blockchain of Learning Logs (BOLL) system (Ocheja, Flanagan, Ueda, et al., 2019) that connects learning behaviour logs and digital contents across institutes. We extend the BOLL system to include assessment results by integrating scores reporting, blockchain encoding, decentralized analytics opt-in/out function and a visualization to support data-driven decision making by stakeholders. To highlight the potential impact of our proposal, we present a simple orchestration of how our proposed framework can be used in high school to report assessment results and support students in making college enrolment decisions.

3.5.2 Architecture design

We propose a framework that allow institutes to report assessment results of their students on a decentralized network with strict privacy preservation and learner control as illustrated in Fig. 3.9. When students interact with learning platforms, the learning behaviour logs as well as the outcome of assessments are reported on the blockchain through the BOLL system. In the next subsections, we will discuss the layers in proposed architecture.

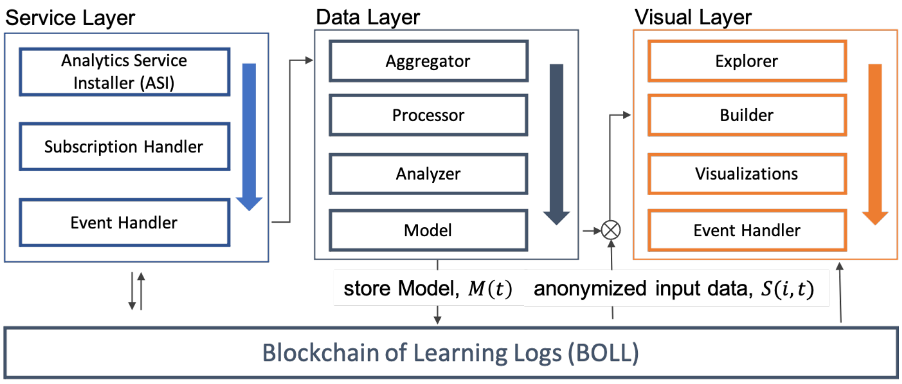


Figure 3.9: Proposed architecture

A. Service layer

This consist of 3 components: Analytics Service Installer (ASI), Subscription Handler and an Event Handler.

Analytics Service Installer (ASI): The Analytics Service Installer (ASI) within the service layer allow researchers or learning service providers to install a Learning Analytics Service (LAS) that can aid students' learning objectives such as predictions, interventions and recommendations. Each LAS is implemented in form of a smart contract and must specify the category of the service provided, the student data required, opt-in/opt-out functions, and a compute function for obtaining results.

Subscription Handler: The subscription handler is a set of smart contracts and utilities that help learners manage their subscription to LAS's. The learners can subscribe to a LAS by granting the service provider access to the necessary data required by the LAS. Upon subscription, The LAS can then retrieve the corresponding learner's data from various institutes connected on the BOLL system. These data is then used by the LAS to build specific visualizations or outputs as specified in the LAS's compute function.

Event Handler: The event handler enable data and visualization layers to stay updated when data associated with LAS's change and may require updates to the previous models or visualizations. This is achieved by having specific smart contracts that are triggered whenever learners subscribe or unsubscribe from a service, and new data is provided or excluded such that it notably affects the previously computed model's performance.

B. Data layer

The data layer is made up of the aggregator, data processors, analyzers and the resulting model.

Aggregator: The aggregator helps to retrieve the data of all the subscribers to a given LAS. A typical implementation consist of checking LAS smart contract for the list of subscribers, the required data and submitting a signed request to retrieve the relevant data from each institute attended by the student. The signed request provides a location for the requested data to be submitted to and ensures that a time signature exist for each data retrieved by the LAS provider. Once the signed request is verified, each concerned institute submits the requested data to the specified location. *Processor and Analyzer:* These components provide similar functions as the data processing and analysis tasks in data mining. The processor typically consist of data validation, cleaning, feature extraction and engineering. The Analyzer handles data analysis by resolving the optimal parameters for the best model. It is important to mention that the processor and analyzer is unique for each LAS. *Model:* This is the final model resulting from the data provided at a specific time, t. A LAS can provide an initial model to prevent a cold-start problem for early subscribers. In such cases, the LAS is expected to have a prior awareness of what kind of data the subscribers may provide as input to the model. In a situation where the LAS does not provide an initial model, the service may build the model as more subscribers opt-in to the service and provide their data for building the model.

C. Visual layer

The visual layer is directly user facing and provide interfaces for querying various insights deduced from the previously provided data. The Explorer allow subscribers to invoke the compute functions implemented by LAS's. Based on the type of exploration or compute function selected, the Builder invokes the relevant LAS's compute function; providing the subscriber's details as input. The compute function uses the provided data and the existing model (if any) to provide outputs that can be displayed on the visualizations. The type of visualization displayed is based on the value returned by the compute function for example, series data would result in a time series or scatter plot visualization.

3.5.3 Smart contracts schema

When enabling decentralized access to user data and use of such data for analytics, it is important to consider the security and privacy of the stakeholders. For our proposed decentralized learning analytics platform, we propose some smart contracts and functions that are necessary for realizing the analytics and privacy requirements of stakeholders. In this section, we will address policies such as modifications to existing smart contracts to take into account assessment data, extendable templates for LAS smart contracts, and other smart contracts that specify what rules apply to LAS smart contracts based on users' subscriptions.

A. LLPC and assessment data

Recall that in Section 3.2.3 we introduced various smart contracts that enable the connection and transfer of learning logs. In that section, we introduced the LLPC which represents a proof of existence of a learner's learning data on a provider's platform. This smart contract is important to our implementation of LAS and decentralized learning analytics in general. The reason is that the LLPC contain the permission and data required for learning analytics. Thus, the LLPC smart contract must be consulted before any learning behaviour logs can be accessed. The LLPC already contain assessment data but not as metadata. The availability of metadata would make learning analytics much easier but as information in the LLPC are particular to a learner, we avoid such exposure through metadata and propose a different approach. Our proposal here is that for all LAS that require access to such data, a separate smart contract should be designed and that contract can then request permission to access the LLPC's data and be listed as a permitted accessor in the *permissions* array of the LLPC. In the next section, we discuss the structure of the LAS policy and service smart contract.

B. LAS policy smart contract

The LAS policy smart contract provides some utilities for service providers to describe certain characteristics of the LAS such as required data, subscriber info, subscription expiration, and access logging. In Table 3.5 we define the basic attributes of the LAS policy smart contract. The *user* attribute points to the service provider while the *subscriber* attribute contains the blockchain address of the user. The *URI* is an address that uniquely identifies this type of policy contract and it is specific to a give LAS. The *accessibleURIs* holds a list of all user information types that the LAS for which this policy is issued would like to access. This could be URI's pointing to information such as assessment results and behaviour logs. The *logAccess* function is provided to enable logging of access to user information. This function is invoked each time the corresponding subscriber information listed in the *accessibleURIs* is accessed by the LAS. Thus, users can always know when their data is being used by the service provider and can trace such usage.

Table 3.5: Learning Analytics Service (LAS) Policy Smart Contract - LASPC

Attribute	Description
owner	address of LAS provider
subscriber	address of the LAS user
URI	unique identifier for this policy
accessible URIs	unique identifiers of all user data types accessible to this LAS
expiration	expiry date of this subscription
log access	creates a footprint anytime user data is accessed by this LAS

C. LAS smart contract

The LAS smart contract provides a guiding template for service providers to extend and orchestrate LAS agreements with learners. In Table 3.6, we describe the attributes of this base smart contract. A LAS smart contract must specify which *policy* should be applicable. When a learner subscribes to a service, they must agree to the accompanying policy and agree to share the required information. Upon subscription, the LAS smart contract is registered as an accessor to the corresponding LLPC's where the require data for analytics is stored. Each time a learner's data is read from the LLPC by a LAS smart contract, the *logAccess* function of the associate policy is invoked and a footprint is left on the blockchain for audit trace. The *unsubscribe* function provides a way for learners to opt out of a LAS. When the *unsubscribe* function is invoked, the LAS smart contract is unregistered from the LLPC and the learner is removed from the subscribers list.

Table 3.6: Learning Analytics Service (LAS) Smart Contract - LASC

Attribute	**Description**
owner	address of LAS provider
subscribers	addresses of all users of this LAS
URI	unique identifier for this LAS
policy	the URI of the LASPC associated with this LAS
subscribe	a function called by a user who wants to subscribe to this LAS
unsubscribe	a function called by a user who wants to unsubscribe from this LAS

Chapter 4

Experiments

4.1 Managing lifelong learning records through the blockchain

4.1.1 Aim and research questions

To provide a concrete implementation of a blockchain-based platform for learning logs based on previous research by Ocheja et al., 2018. We show that it is possible to achieve a privacy-preserving lifelong learning log using the blockchain with defined smart contracts, discuss resource requirements, and the benefits of our proposed system. We also discuss potential challenges that may be faced and provide solutions on how such issues could be tackled. There are two main research questions in this study:

1. How can we implement a privacy-preserving connection of lifelong learning logs on the blockchain?
2. What are the resources requirements and implications of connecting lifelong learning logs on the blockchain?

4.1.2 Methodology

In carrying out this research, we adopted the Design-Based Research (DBR) methodology(Wang & Hannafin, 2005). Wang and Hannafin (2005) defined DBR as a research method which focuses on exploring systematic but flexible techniques targeted at improving educational practices through iterative analysis, design, development, and implementation requiring collaboration between researchers and practitioners and leading to new useful principles. The idea of iterative analysis as applied in DBR, helps to validate design decisions. In the event that a design approach fails in the validation phase with real-world practitioners, another cycle of iteration can consider alternative techniques. Such a repetitive task makes it possible to arrive at a more feasible implementation that meet the needs of the end-users.

In the design of our proposed system, we first conducted a literature review on previous works that have attempted to enable lifelong learning logs. Specifically, we used the framework proposed by Ocheja et al. (2018) as a guide in deciding how functionalities on our proposed system are different from other systems. We also considered the different stakeholders that are involved in managing and accessing learning records, such as learners, teachers, administrators, researchers and other third parties so as to ensure that our system caters for their needs. This was carried out by observing current processes and concerns in academic institutions involving these stakeholders, such as privacy, security, accessibility, availability, and consistency of learning records.

Consequently, we developed smart contracts that reflect how learning records are generated and how access to them is controlled and managed. As learning records are categorized by action words or verbs from which they resulted from (Consortium, 2013; Learning, 2016b), we adopted an action verb-based method of storing and managing privacy of learning records on our proposed system. In this case, learning records of the same action verb for a particular learner, are written to the same smart contract on our proposed system alongside their permissions.

We validated our design by using data from learning tools in our institute's production environment. These data contain information about learners' activities on the learning tools including quiz, read, assignments, view and other events. To validate our design using this data, we developed scripts that simulate the creation of these learning events. The output of each simulated event is then written to our proposed system. This simulation approach of validating our design is useful in this work as most features of our proposed system can be programmatically triggered.

In this experiment, we also measure the performance of the BOLL system. To carry out this experiment, it is required to have at least a setup as shown in Figure 3.4. We ran this setup on a Dell EMC PowerEdge R530 Hardware (16GB RAM, 512 SSD) with Ubuntu 16.04 Server installed. Also, we setup two other similar instances of Geth in Figure 3.4 on the same server so as to ensure distributed mining of transactions.

Our key performance indicators for the BOLL system specifically considers the amount of computational resources required to mine: intermediate transactions, write, update and access learning records. To measure these, we used the gas usage and timestamp parameter of each transaction to understand both computational and time resource requirements. Using learning records generated by students' activities on Moodle LMS and BookRoll, we simulated some of the processes outlined in Figure 3.6. Learning records are generated and logged on the OpenLRW whenever students use BookRoll. Table 4.1 shows the numerical description of the population and sample space. 651 students generated 498,842 records of which 291 students' learning records reflected in the randomly sampled 500 learning records to be written on the blockchain.

Table 4.1: Test data description.

	Number of Learning Records	**Number of Users**	**Number of Action Verbs**
Total	498,842	651	8
Sampled	500	291	7

Writing these learning records on the blockchain require creating, updating and validating different smart contracts as shown in the process outlined in Figure 3.6. The distribution of transactions generated as a result of the various operations required to write 500 learning records of 291 students is shown in Figure 4.1. A total of 3,104 transactions were generated with 1,000 of them coming from permissions and indexing operations (UIC and PIC) on the learning records.

4.1.3 Results

From our test, we observed that processing various smart contracts on the BOLL system requires different computational cost. In Table 4.2, we show a list of these transactions. As stated earlier, gas usage is a representation of the complexity of an operation. Currently, there are no standards on how to determine the equivalent conversion from gas to physical currency. Some factors could guide the determination of such including the cost of electricity, servers and maybe cost of labour. In our implementation, we note that while create operations (1, 3, and 5 in Table 4.2) require more gas usage, update operations are less expensive (2, 4 and 6 in Table 4.2). Creating an LLPC is computationally complex and requires 1,814,374 gas to process. This is because the permissions and learning records indexing strategy are defined in this smart contract and installed upon creation. Similarly, PIC and UIC require 1,030,138 gas to process because of the indexing strategy defined in the index smart contracts.

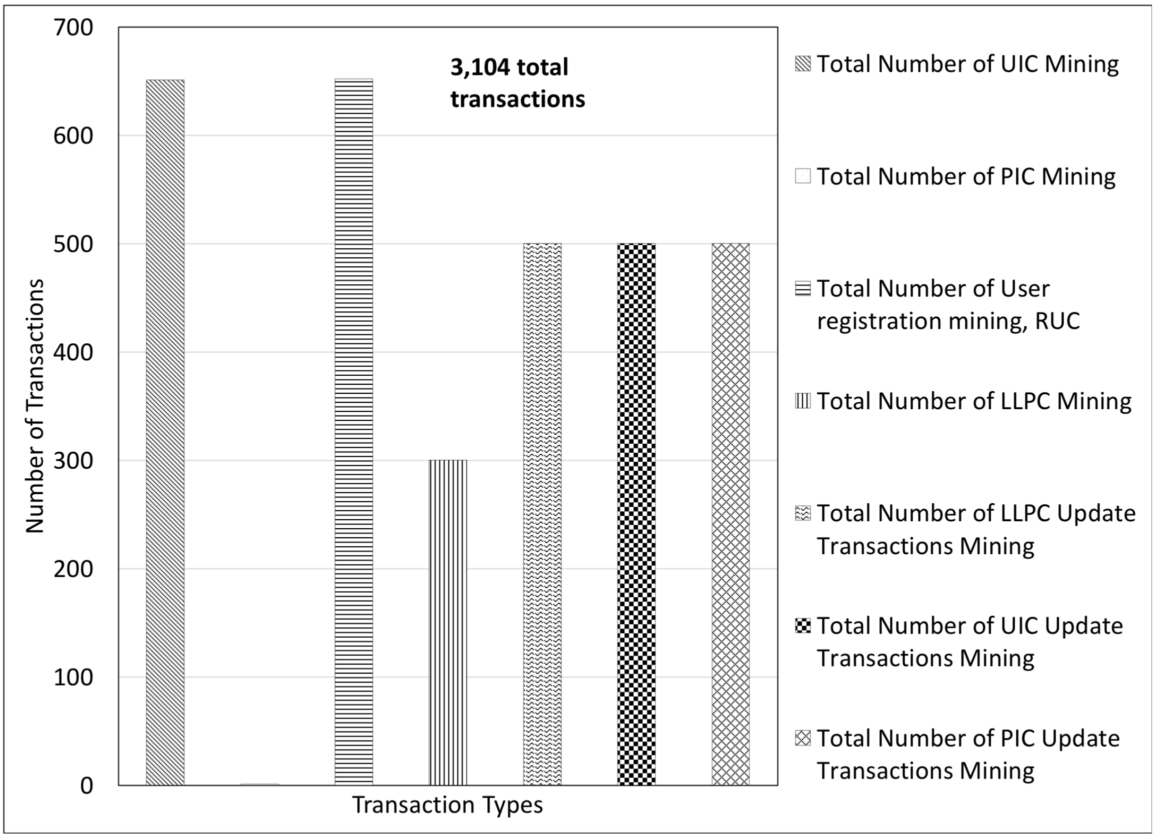


Figure 4.1: Smart contracts mining operations.

Table 4.2: Computational cost of smart contract operations.

S/No.	**Smart Contract**	**Action**	**Frequency of Operations**	**Average Cost (gas 10^3)**
1.	PIC	Create	1-time	1,030
2.	PIC	Update	Every time	115
3.	UIC	Create	1-time	1,030
4.	UIC	Update	Every time	27
5.	LLPC	Create	On new action verb	1,814
6.	LLPC	Update	Every time	298
7.	RUC	Update	On user registration	55

In our test case, we obtained a waiting time, W_t of 14 minutes per transaction. Importantly, W_t is different from the time it takes to mine a transaction. On the Ethereum blockchain, this is a function of the current complexity of the Proof of Work otherwise referred to as the difficulty. The Proof of Work (PoW) is a cryptographic puzzle that involves finding a value whose SHA-256 hash begins with a given number of zero bits. This is enforced to ensure that mining nodes on the blockchain have done some amount of work and the resulting write operations were done in consensus with other participants on the network agreeing to the result of the PoW. It also makes revocation of write operations difficult.

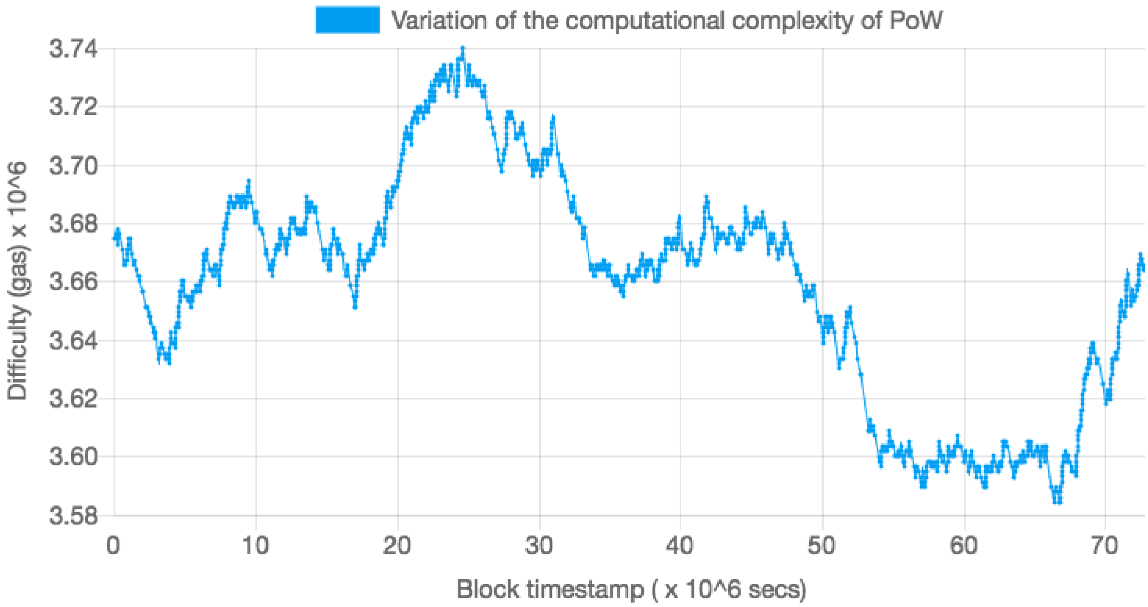


Figure 4.2: PoW computational complexity over time on BOLL.

In Figure 4.2, we show a plot of difficulty in mining the different blocks representing our learning log transactions over time. The difficulty increases or decreases depending on the amount of computational resources available and the computational power spent on computing the preceding puzzle across the system. In Figure 4.3, we also show a plot of the time elapsed between transaction creation and its effective mining over the different blocks' timestamp. The graph shows a near linear increase in time difference because transactions are mined in turns hence our earlier calculation of a 14-minute waiting time.

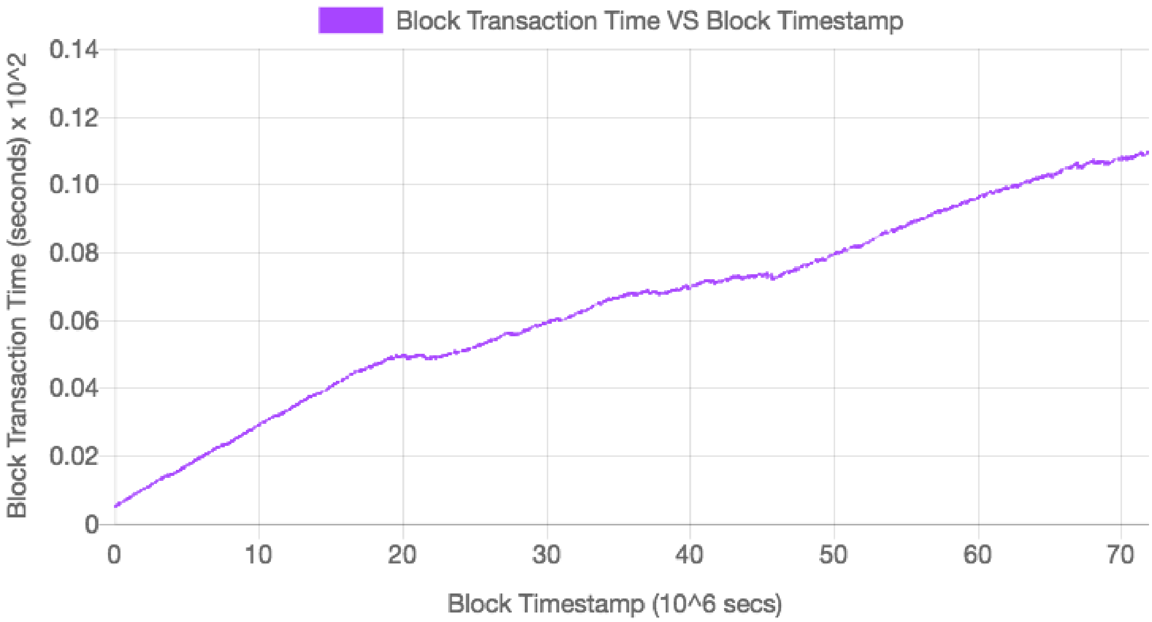


Figure 4.3: Time elapsed between transaction submission and mining completion vs mining completion time.

Table 4.3: Comparison of BOLL system to other learning infrastructure.

Learning Infrastructure	**Single Sign-On (SSO)**	**Connected Learning Logs**	**Fast Write of Bulk Records**	**Decentralized (privacy, security, etc.)**
LMS & LRS with xAPI & Caliper Integration	✓	X	✓	X
IMS CLR	✓	✓	✓	X
BOLL System	✓	✓	△	✓

In Table 4.3, we compare the features and performance of our BOLL System to other learning infrastructure. While most learning infrastructure provide support for Single-Sign-On (SSO), only IMS CLR and BOLL provide support for connecting learning logs. Consequently, systems that do not provide support for connecting learning logs often face the cold start problem. However, only BOLL system offers a high degree of privacy through smart contracts-based access authorization where learners can actively determine who can collect their learning logs and access them at a later time.

4.1.4 Discussion

Here we discuss some discoveries, questions and problems that arose during the implementation and testing of the proposed BOLL system.

A. Privacy

On BOLL, the identity of learners from their institution's learning tools is not shared between different institutions. Instead, we generate a sequence of bytes called address for each learner upon registration. For subsequent record look-ups, we use this address as a way of tracking their records on BOLL. This design ensures that only authorized parties can link records on BOLL to the right learner.

From our implementation, we confirmed that unless one has access to the learner's private key, it is impossible to access their learning records without their permission. This is made possible by the inherent security of the blockchain, installed smart contracts and given that all access to such learning records are made through BOLL. We make the assumption that learners would guard their private key from unauthorized access. A 3rd party can have read, write or grant privileges to the LLPC smart contract containing a learner's learning records. By default, only the learner and their institution where such learning records were generated can grant access to 3rd parties. If a 3rd party requires access to these learning records, they can send an access request to the learner. The learner or their institution can then choose whether to grant any or all of the three access privileges to the requesting 3rd party.

However, we observed a limitation in using action verb-based smart contracts. In grouping learning records according to action verbs, if a learner gives an institution access to read one action verb, such institution is authorized to read all their learning records having that action verb regardless of the learning material from which the learning event emanated. This problem can be solved by extracting identifiers for different learning materials and use a pair of these identifiers and action verbs as way of keeping access to learning records limited to learning materials.

Also, a learner may choose to deauthorize a 3rd party from having access to their learning records. This is possible by removing the 3rd party's address from the list of authorized accessors in the LLPC. However, we do not currently allow the deauthorization of the learner's current institution especially when no other institution has access to their learning records. This is because if all institutions are deauthorized from accessing a learner's records, it will be impossible to locate their records on the institution's platform. We address this issue while discussing *'demise of an institution'* and we suggest that prior to such deauthorization, a learner should enable backup of their data to an authorized data storage site on BOLL.

B. Performance

Our BOLL system currently has an average waiting time of 14 minutes. This means that for a new learning log to be written on the BOLL network (writing operation may include learner registration, LLPC contract creation and/or updating and indexing), one would have to wait for an average of 14 minutes. This time might be acceptable for some use case where learning logs are not required to be read from the blockchain in real-time as soon as they are generated on the learning platforms. In fact, W_t can be much less than 14-minutes, an isolated case where LLPC is only being updated, it would take between 30 seconds to 2 minutes. We are also currently considering integrating new patch-set from the Ethereum lightning network; an off-chain scalable solution that in some sense allows for distributed and faster mining.

C. Installable smart contracts

We have defined a number of installable smart contracts for decentralized control and access of learning logs: RLPC, UIC, PIC, LLPC. While permissions may differ for different types of learning logs and users, our implementation considers a generic permission structure for all learning logs. We also treat access authorization in a similar manner but empower the users with the ability to grant or revoke access at any time using pre-installed smart contracts. We consider it interesting to look at the various scenarios that might occur when learning logs are of different types and governed by different data policies. One possible solution would be the presentation of smart contracts in a form where learners can understand the concept of the smart contract, and be able to select an appropriate smart contract that may suit their needs from an open pool of personal learning logs smart contracts.

D. Demise of an institution

As only a hash of the learning log and its location is recorded on the blockchain, there is a possibility of a learning log outliving its host institution. For example, a student might graduate from an institution and 10 years later, that institution ceases to exist. In a case where all computing facilities such as the LRS of that institution is also shutdown, then the learning logs whose references are held on the blockchain cannot be retrieved anymore. To solve this problem, we envisage a learning blockchain where not just institutions exist on the network but also third parties who can offer data backup services. These third parties do not act as mediators in anyway but rather serve as storage centers for learners on the blockchain. Another alternative will be to specify smart contract policies where learning records are held on file for a certain duration of time. Currently, we do not recommend that the blockchain should replace traditional databases except for simple-size data.

E. Cost

Cost of computation and infrastructure are the key factors in determining the budget for a learning blockchain. In our implementation of the BOLL system, we incurred some cost in procuring and setting up the servers on which the blockchain node was hosted, electricity bills, internet, etc. In deciding how miners on the learning blockchain get rewarded, these costs need to be factored in. Whether such cost is transferred to the learners or institutions is an open question for stakeholders. Whichever might be decided, the blockchain provides a way to measure such cost through gas usage.

4.2 Visualization of education blockchain data

4.2.1 Aim and research questions

To understand the expectations of stakeholders that use the information provided by education blockchain systems, we modeled this problem as a typical case of visualizing distributed academic records and conducted a qualitative inquiry (Patton et al., 1980). Before conducting the interviews, we first presented the situation of disconnected learning records to the interviewees and demonstrated how such problems could be solved through the blockchain. The flow of the interview is shown in figure 4.4. Our focus user group include two Mathematics teachers each from a Junior High School (JHS) and High School

(HS) in Japan. We first demonstrated to the teachers how it is possible to connect academic records of students through the blockchain using the BOLL system (Ocheja, Flanagan, Ueda, et al., 2019). Specifically, in the previous year, we used the BOLL system to collect learning logs of current HS grade 1 students while they were enrolled in JHS grade 3 Mathematics. It is important to note that the HS and JHS are different schools, students' accounts in each school are different and no connection exist between the two schools to enable data connection and/or transfer. This disconnection makes it necessary to develop a system like BOLL that can connect the records of students that move from the JHS to the HS. Also, since JHS grade 3 Mathematics is a prerequisite to HS grade 1 Mathematics, it becomes suitable to provide access to its contents for revisions, and/or reflections.

4.2.2 Methodology

The flow of the interview is shown in figure 4.4. Our focus user group include two Mathematics teachers each from a Junior High School (JHS) and High School (HS) in Japan. We first demonstrated to the teachers how it is possible to connect academic records of students through the blockchain using the BOLL system (Ocheja, Flanagan, Ueda, et al., 2019). Specifically, in the previous year, we used the BOLL system to collect learning logs of current HS grade 1 students while they were enrolled in JHS grade 3 Mathematics. It is important to note that the HS and JHS are different schools, students' accounts in each school are different and no connection exist between the two schools to enable data connection and/or transfer. This disconnection makes it necessary to develop a system like BOLL that can connect the records of students that move from the JHS to the HS. Also, since JHS grade 3 Mathematics is a prerequisite to HS grade 1 Mathematics, it becomes suitable to provide access to its contents for revisions, and/or reflections.

Our primary method of evaluating our proposal includes conducting a stakeholder relevance assessment. This is done by conducting semi-structured interviews Longhurst, 2003 which is a common qualitative evaluation method for blockchain research Toufaily et al., 2021. The interview was conducted with four teachers with selected questions bothering on: the relevance of past learning records and visualizations, situations that necessitate such visualizations, current alternatives, important aspects of prior learning data and the potential impact when access to prior learner data is possible.

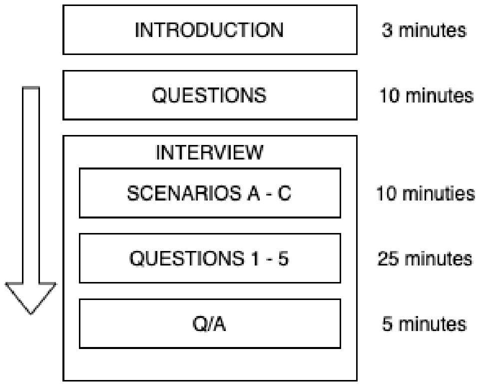


Figure 4.4: Interview flow

4.2.3 Results

We present below the results of the qualitative inquiry about our proposals with four teachers. It is important to mention that the results from these interviews have been used in the design of visualizations for education blockchain as earlier presented in Section 3.4.2 of this research.

A. How important is it to access past learning records?

Four out of four teachers agreed that access to past learning records is important. Some of the factors that necessitate access to past learning records, as mentioned by the teachers include: making informed class designs, enabling personalized learning and provides a new way to assess students' ability. However, one of the teachers noted that the current deployment of learning tools within their institute is yet to collect enough data from multiple learning environments such as cram schools which is very popular among students in this region. We acknowledge that this could be a challenge especially for early adopters of the education blockchain. Thus, a probable solution to this problem is the provision of a framework that inter-operates seamlessly with existing learning tools.

B. In what type of scenarios is access to past learning records considered useful? Give examples.

The interviewees gave many examples of situations where access to past learning records is desirable. For example, when students change class either within the same school or across schools or across different levels (e.g., Junior High School to High School), access to past learning records will be useful in understanding the learner's ability and attitude towards learning. This will help the teachers to provide personalized contents that meet the needs and challenges of their students. This corroborates with our earlier statement about not having enough data to provide personalization from the beginning of a class: the *cold-start problem*. However, one teacher mentioned that even when past learning records are accessible, there is a possibility of not finding any new information. While this may be true, what transcripts lack is the ability to state the reasons a student might have obtained poorer or better grades. When access to past learning records is possible, we can find additional information by using learning analytics to measure. This standpoint is re-echoed by another teacher who mentioned that access to past learning records can provide engagement information which in turn can reflect how committed students are to their studies. Therefore, an education blockchain data visualization that meets the various scenarios identified by these teachers is desirable.

C. How do teachers currently fill the gap when they cannot access past learning records?

All four teachers acknowledged that they conduct some form of inquiry to get additional information about students' past learning or current academic situation. For example, all the teachers that were interviewed mentioned that they consult with the teachers of their students in other courses or previous level and in some cases, teachers administer quizzes to assess what the students can recall from past learning. One teacher also mentioned that to get information about students' past learning, some teachers may ask the students which cram school they attended. This is because the teachers believe that the cram school a student attends can influence what they know or how well they have been taught. Our observation here is that there is a continuous inquisition for more information: the teachers want to know more about their students so as to provide precise intervention. Thus, access to academic records of these students beyond one institution is important.

D. What information do teachers consider useful in students' past records?

From the responses collected during our interviews, it is evident that teachers consider access to students' past learning records to be important. Two of the teachers were interested in accessing students' engagement information such as the learning materials students have used, how frequent the students used such learning contents, and the type of quiz questions students attempted (easy or difficulty). One of the teachers also mentioned that information on students' grammatical knowledge will also be useful. Here, we notice that important academic credentials desired by the teachers go beyond certificates and transcripts. Our review of previous implementations of education blockchains do not address access to information such as students' engagement and learning materials used. Hence, it becomes necessary to include this type of information in future implementations of education blockchain visualization.

E. What are the expectations or projected outcome when teachers can access past records of their students?

One of the teachers envisaged that with access to past learning records, teachers will be able to provide personalized contents for their students which could lead to better understanding, ease of learning and improved performance. Another teacher also noted that when past learning records can be accessed before teaching starts, personalized teaching can be implemented from the start of the class. While one teacher could not imagine what the possible impact could be, they were positive about using such a system in their class. Our observation here is that the teachers are optimistic about what can be achieved when access to students' prior learning data is possible. We share in the views of these teachers ranging from solving the cold-start problem to enabling provision of personalized learning contents.

To address the above needs of the teachers, we first reviewed the features and data connected through the education blockchain such as the Boll system. Next we designed and implemented visualizations that can improve the data awareness of stakeholders. An important question that we must first answer is this: What does blockchain bring to education that makes it so peculiar in comparison with current learning systems and platform? Education blockchain mainly brings the features of the blockchain to the education space. These features include: decentralization, sharing of data across multiple systems, traceability, privacy, and security. For each of these features, we will discuss how they affect education systems and visualizations on LAD's. In the subsections to follow, we will discuss the above key elements of our proposed visualization framework.

4.2.4 Decentralization and data sharing

This primarily means that multiple parties can communicate, make decisions and exchange information in a trustless manner. In education, one may argue that trust might not be a big problem but this is not true. Currently, most countries regulate academic institutions and provide rules to guide how they operate. This is necessary for reasons including: standardization, fidelity, quality and verification of academic processes. The blockchain technology can facilitate all these features in a democratic manner through its decentralization feature. In simple terms, institutions can join a common blockchain where the rules for standardized processes, dedication to set goals, academic quality and verification of testimonials can be decided by all the members of the network (e.g. public blockchain), a selected group (e.g. consortium blockchain), or a single authority (e.g. private blockchain).

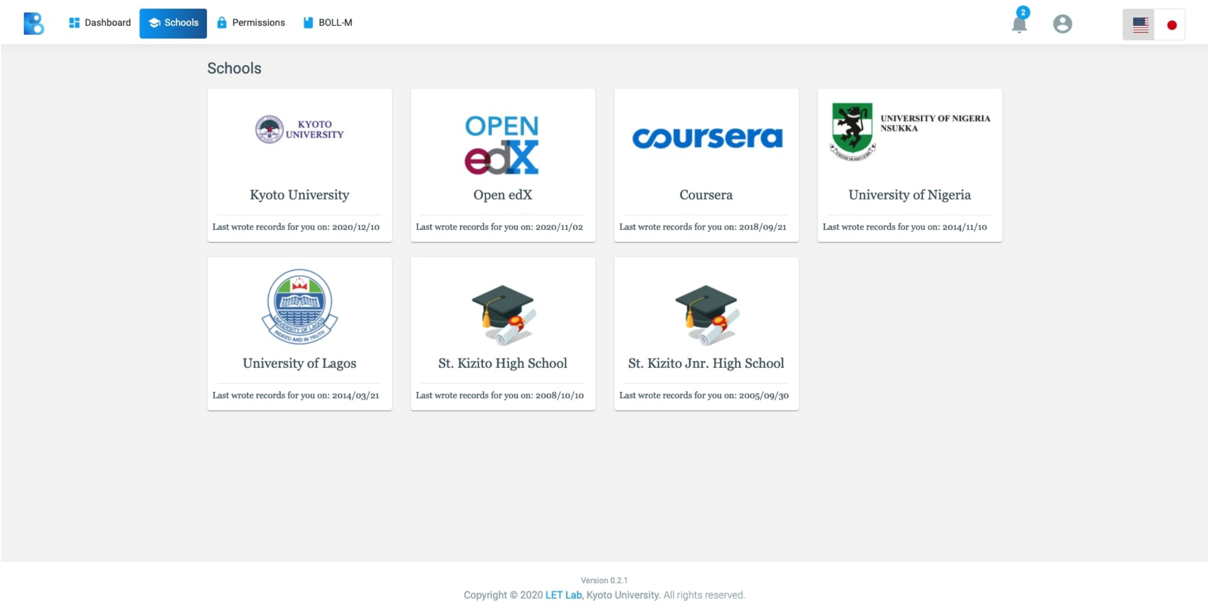


Figure 4.5: Visualization of schools attended.

A. Stakeholders

The parties that are involved and benefit from the decentralization feature of the blockchain include: students, teachers, academic institutions and other organizations. Students would be able to navigate various institutions that are members of the blockchain without worrying about records management separately. Teachers would also be able to navigate institutes more seamlessly, collaborate and access history data of their students across the different institutes they have attended so as to be able to support them better (personalized learning, competence depth checking, etc.). Academic institutions and their staff can handle administrative matters quickly especially during enrollment, transcripts issuance, records transfer, teaching and learning outcome analysis. As for other organizations, it opens up the academic environment for more partnership to deliver more learning tools for advancing students learning and enable a much faster way to exchange data between formal academic processes and informal education or learning companies and other companies in general.

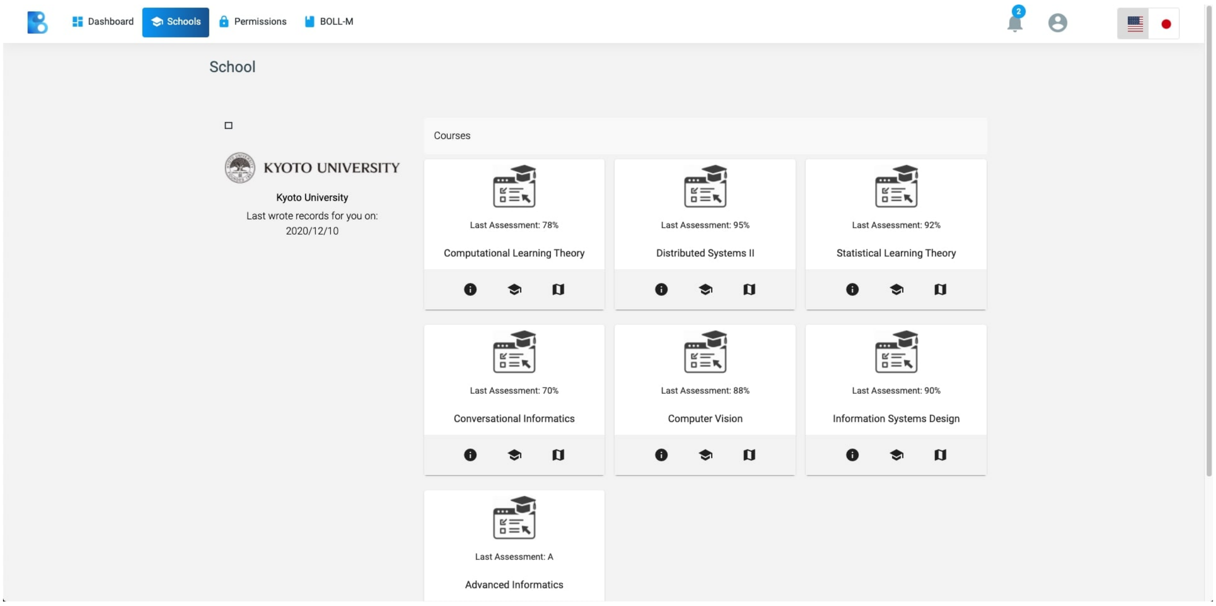


Figure 4.6: Visualization of courses at each school.

B. Blockchain data visualization

To reflect decentralization paradigm on LAD's, it will be important to make available visualizations for students to access their data across multiple schools. For example, in figure 4.5, we show the BOLL visualization of a student's past schools attended. This student has attended 5 different schools and has taken courses on 2 different MOOC platforms. Each of these schools have connected to the BOLL platform and have independently reported the student's records on the blockchain with the student's consent. On click on any of these learning institutes, the course(s) the student enrolled at that school will be displayed as shown in figure 4.6. Each of the courses contain additional information including: the enrolment information, student's learning logs comprising of

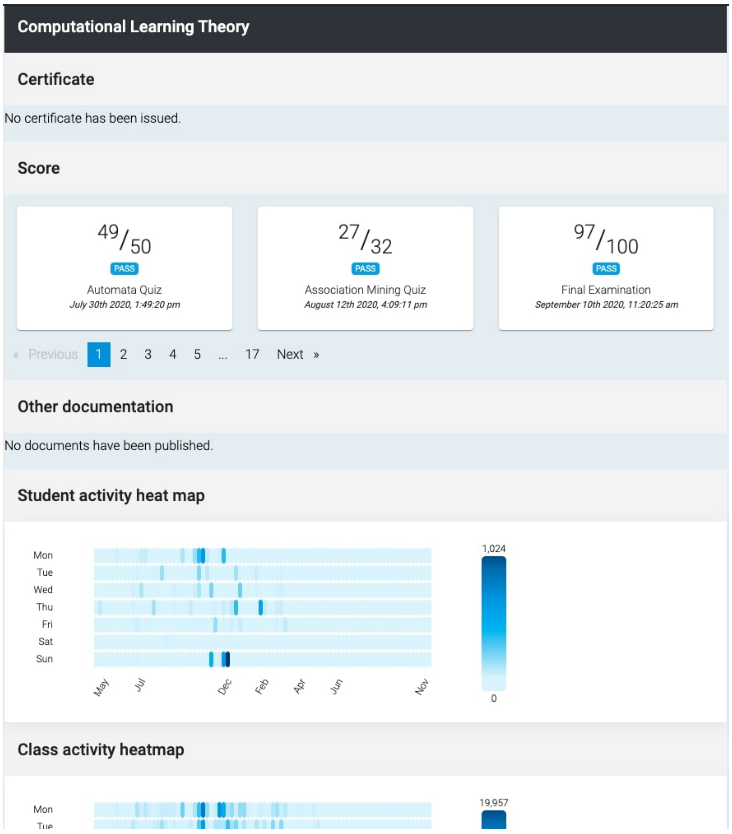


Figure 4.7: Visualization of data in each course.

engagement information, scores, and grades, as shown in figure 4.7. Other information available in the student's record on the blockchain include learning materials used, and the knowledge map information for specific learning materials (if provided).

In our proposed visualization, we use the knowledge map concept proposed by Flanagan et al., 2019 to show what concepts students have covered and to what extent as show in figure 4.8. Each of the nodes on the graph represent concepts in a given course or learning material. The green dial on each node indicates how well the students have master each concept (0 - 100%). This is calculated by using students' performance in quizzes attached to each node, their learning engagement (time spent on each concept) and the weight of each concept on the graph. For a better illustration, figures 4.9 and 4.10 show details of Junior High School (JHS) 2 Math and JHS 3 Math topics. The concepts in figure 4.10 are prerequisites to concepts in figure 4.9. Access to such information can enable teachers to know how prepared students are to learn a new topic as in figure 4.11 and potential challenges. In this work, we use a simple approach where the performances of the students are distributed across 4 cohorts using their score or engagement (x): low-

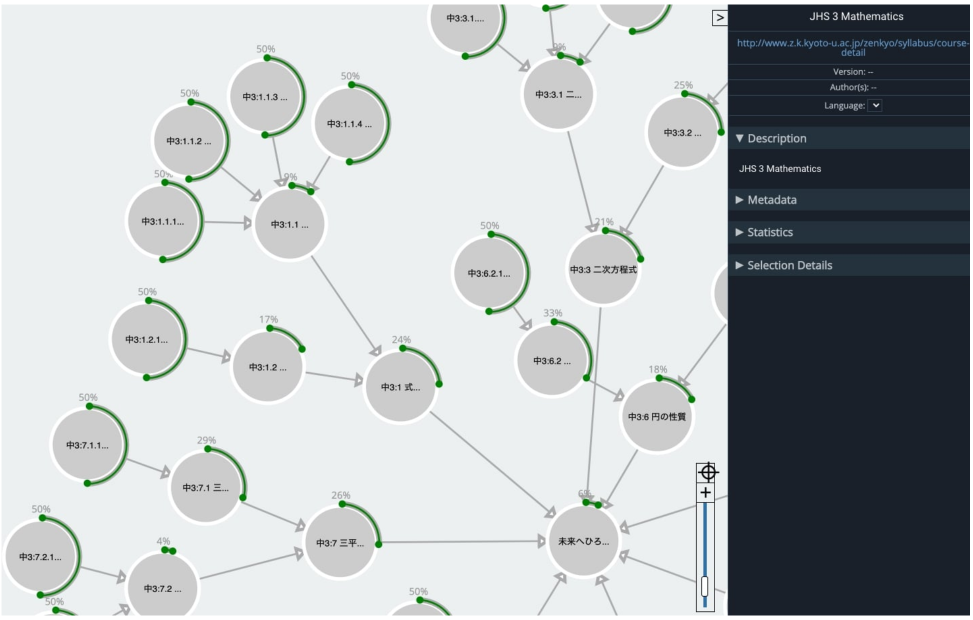


Figure 4.8: Visualization of decentralized data for students.

est ($x \leq 25\%$), low ($25\% < x \leq 50\%$), high ($50\% < x \leq 75\%$) and highest ($75\% < x \leq 100\%$). Teachers can assign tasks to students in these different cohorts. In figure 4.12 we show an example of a visualization for assigning a revision task on a chosen topic to students in a given cohort.

It is important to mention that while the above visualizations may be found in regular learning analytics dashboards, the data used for these type of cross-institute visualizations is made possible through the blockchain. For example, regular analytics dashboards do not have access to students' learning records beyond the institute where the students are currently enrolled. The BOLL system (Ocheja et al., 2018) makes it possible to link academic records of students across multiple schools using the blockchain. Also, access to learning materials used at different learning institutes is also made possible through the blockchain as proposed by Ocheja, Flanagan, and Ogata, 2019.

4.2.5 Privacy

The anonymity of transactions on a blockchain is one of the key attributes that engenders privacy on the network. Members on the blockchain maintain their identity by signing transactions with a pair of public-private key. For education, privacy is a common challenge faced when sharing learning records with third parties. It is important to ensure prioritization of learner's privacy and learners should be in control of their learning

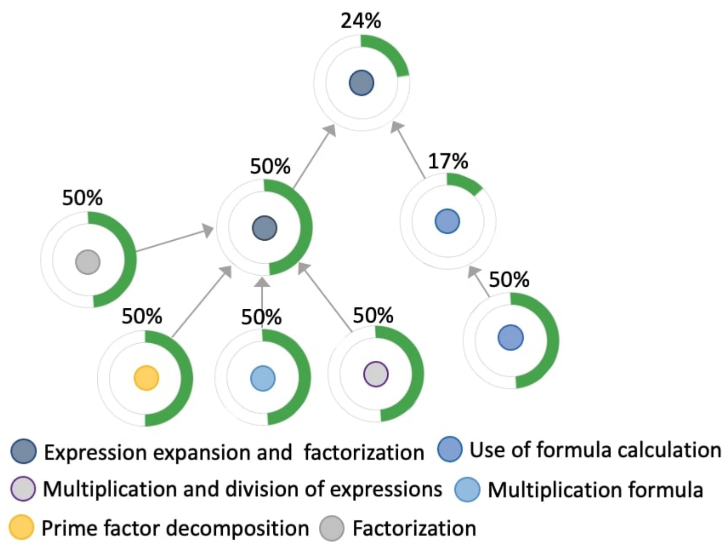


Figure 4.9: Visualization of a student's learning in JHS 3 Math.

data. To achieve privacy through the blockchain, personal information are not stored on the blockchain but rather a reference to data location of such information is held on a block. Whenever a party wants to view such private information, their access rights are first confirmed with the smart contract that protects the requested data. Data owners through smart contracts can enlist parties that can access their data as proposed in (Ocheja, Flanagan, Ueda, et al., 2019). We note that the main privacy feature is facilitated through smart contracts and identity might be revealed on the blockchain after one time permissions access to a record where the address is revealed. However, only hashes on the blockchain can be linked to the identified address and not the actual records.

A. Stakeholders

Students, teachers, institutions and other organizations can benefit from the privacy features of the blockchain. All stakeholders could manage who can access their records or their students' records (for institutions). Students could also further decide whether a member on the blockchain can write records for them or not. Teachers would also be able to request access to their students' records in order to provide teaching support. These stakeholders can also revoke these permissions at anytime.

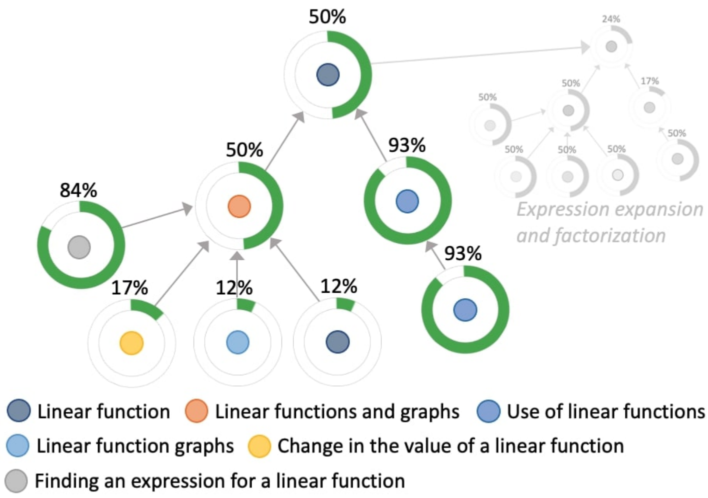


Figure 4.10: Visualization of a student's past learning in JHS 2 Math linked from JHS 3 Math.

B. Blockchain data visualization

Most current LAD's provide a mechanism for managing permissions to learning records. However, these mechanisms are usually done at institutional level and do not yet take into account decentralized systems like the blockchain. In figure 4.13 we show an interface for managing permission to academic records stored on the blockchain. We also show in figure 4.14 an interface for modifying permissions previously granted. Users can use this interface to alter permissions previously granted to other users or third party institutes. Also users may set the expiration date for a permission when granting such permissions. It should be noted that modifying these permissions does not erase records copied to other locations before the permissions were revoked.

4.2.6 Security

In a decentralized system, one key engine is the consensus algorithm. The consensus algorithm stipulates how members of the network reach an agreement before transactions are accepted as valid. These algorithms ensure that whatever action is taken on the network reflects the interest of the majority or at least the interest of the constituted authority. For education blockchain, the security here refers to the ability of the system to withstand attack from malicious entities that seek to alter students' records, inject invalid data or award fake credentials. While these issues are technical, a visualization is important for stakeholders to be aware of network trends, members activities, and overall

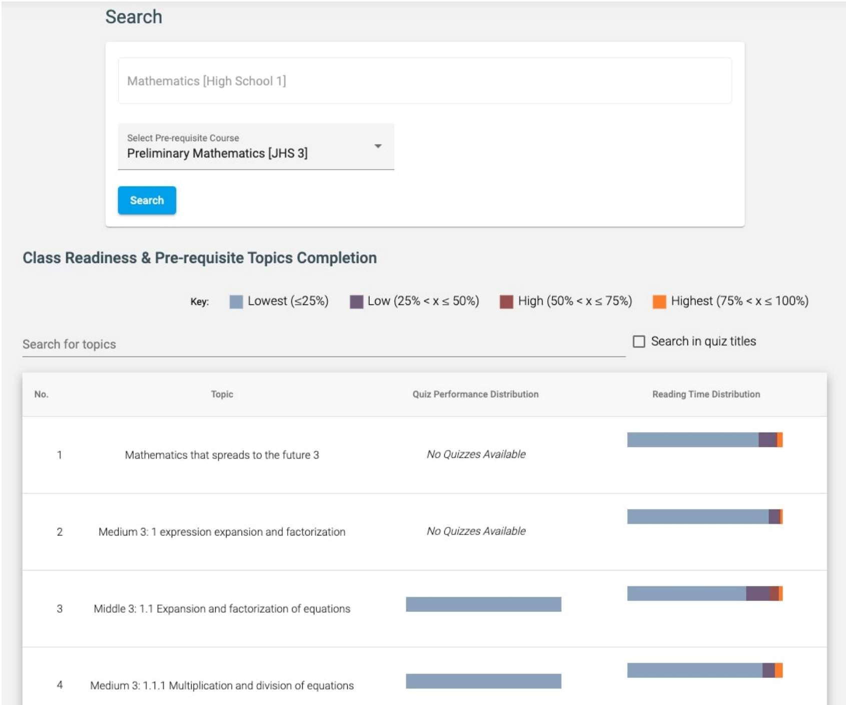


Figure 4.11: Knowledge visualization for past learned concepts and/or prerequisites.

performance of the blockchain.

A. Stakeholders

Institutions and other organizations are the key stakeholders in providing security and ensuring that the network is tamper proof. Although students benefit from a secure education blockchain, we do not recommend a specific visualization on this as this may deviate from their learning or records management objectives. Institutions and other organizations that form primary custodians of academic records could visualize the education blockchain network data including mining operations, nodes enrollment, students' records, transactions processing and computational details.

B. Blockchain data visualization

To visualize security, stakeholders need to be aware of network activities and events that may compromise their assets. Similar to Ethstat (ConsenSys, n.d.), LAD's should provide a visualization for administrators to view network performance, configurations and potential anomalies. In addition, LAD's should also reflect overall network activities of blockchain members to enable institutions understand how various activities are being

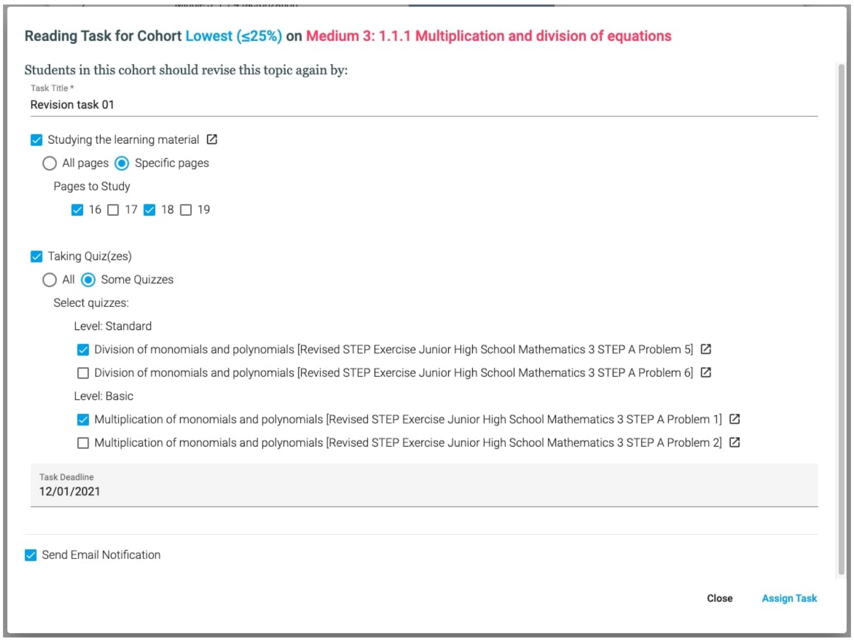


Figure 4.12: Assign a reading task on a specific topic to a selected cohort.

conducted on the network. In figure 4.15 we show an example of such visualization using Ethstat (ConsenSys, n.d.).

However, not all stakeholders can understand the visualization shown on Ethstat due to the lack of knowledge about the blockchain terminologies. Another way to keep stakeholders aware of their security status is through the use of push notification systems (Bell et al., 2011) when a request to access a user's record on a blockchain is issued. This way, users are actively aware of what is going on. In figure 4.16 we show a visualization of how users can setup security notifications regarding various actions that are carried out on their data on the blockchain including: permission requests, approvals, enrolments, records addition, and issuing of credentials. Figure 4.17 show some examples of these types of notifications where the widget name *Messages* refer to messages sent directly to the user by other users or the system (e.g. joining the blockchain, account pairing with new schools, etc.). The *Learning Logs* widget list notifications relating to events on the blockchain that affect the user's academic records such as course enrolment, learning engagement information, issuing of grades and scores, and certificates. The *Permissions* widget list actions relating to authorization requests, grants and denials.

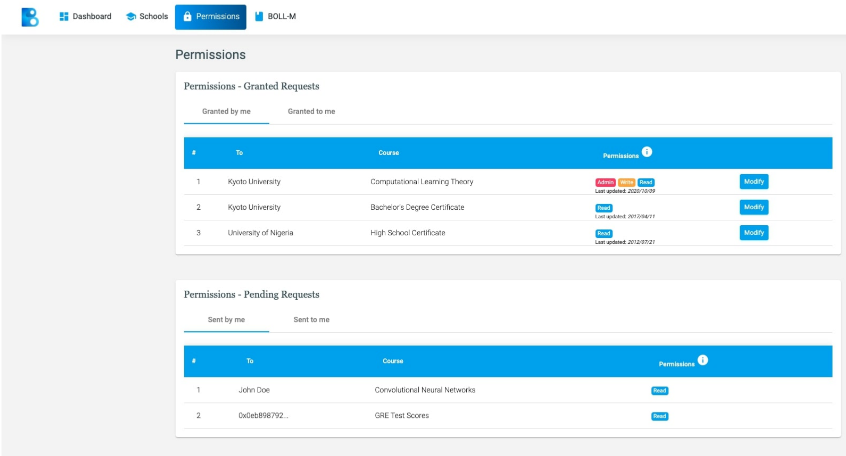


Figure 4.13: Visualization for managing permissions.

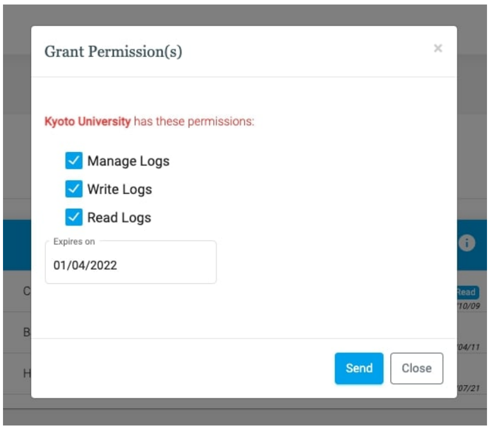


Figure 4.14: Visualization for modifying permissions.

4.2.7 Traceability

The chaining of blocks on the network where each block points to its predecessor makes traceability a core feature of the blockchain. Given any block on the network, it is possible to retrieve all blocks generated before that up to the genesis (first) block on the network. The implication of this for education blockchain can loosely be interpreted as: given an academic record on the network, it is possible to retrieve all other academic records that were issued before the given academic record. This is true if and only if all preceding academic records were also written to the same blockchain network before the given academic record. When such holds, it becomes possible to trace evidences of a learner's learning activities prior to a current learning objective.

Figure 4.15: Visualization of network activity EthstatConsenSys, n.d.

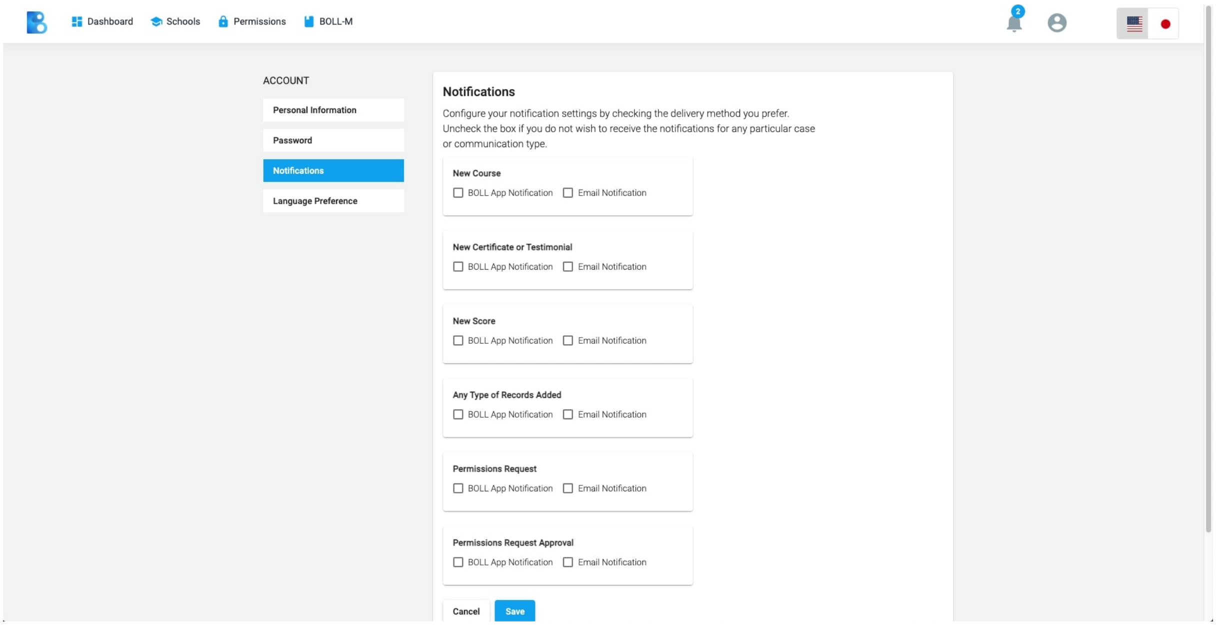


Figure 4.16: Configure security notifications.

A. Stakeholders

The usefulness of traceability is applicable and beneficial to students, teachers, institutions and other organizations. For students, traceability of learning activities provides them with a proof for all their learning engagements especially in cases where it will be naturally difficult to prove that they have undergone such activities. Teachers will also find this useful in diagnosing and supporting students or specific cohorts who find certain topics difficult. In such a situation, teachers may probe into a student's learning history to ascertain what is missing or inadequate in their prerequisite knowledge. Institutions can use such tool to verify competences especially when making decisions on enrollments, hiring, curriculum design or population dynamics.

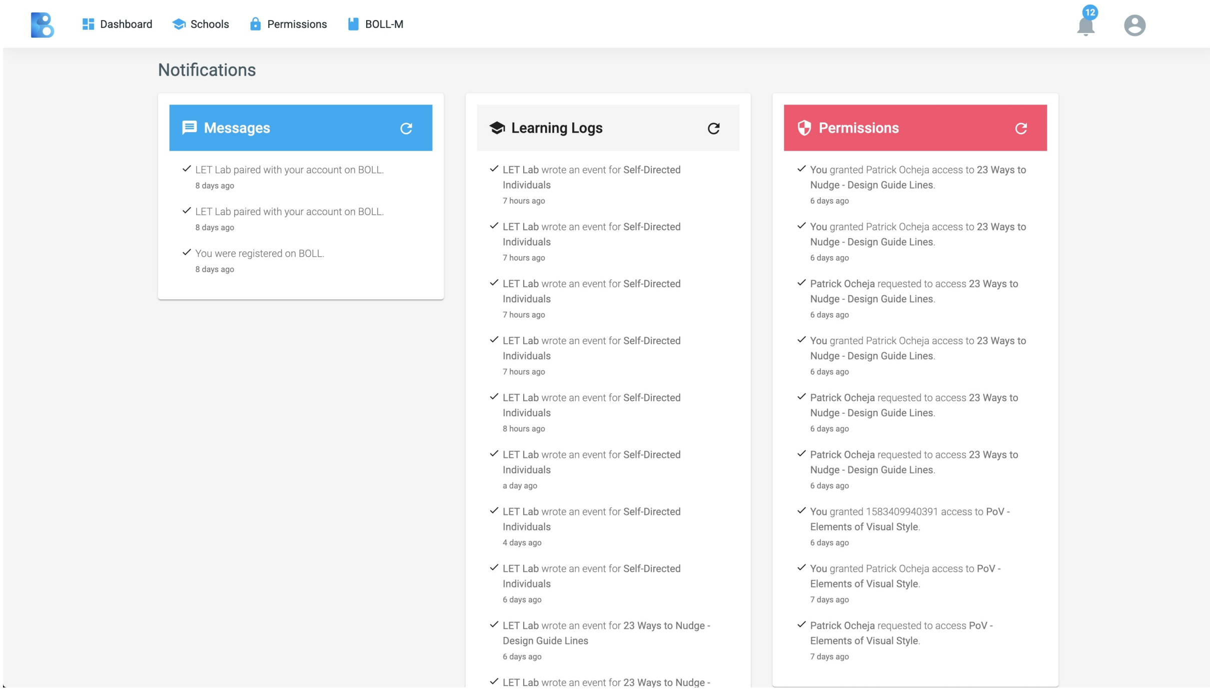


Figure 4.17: Notifications on actions related to a user.

B. Blockchain data visualization

Considering the emergence of blockchain in education, LAD's should be revised to reflect tools for enabling learning traceability. Students should be provided with an interface to interact and access their past learning, how such learning affect current learning activities and a way to revise on such concepts when needed. In figure 4.18 we show an example of how lifelong learning can be connected through education data held on the blockchain as proposed in (Ocheja et al., 2020). This is similar to the visualizations presented in figures 4.8 and 4.11. The difference here is that visualizations for learning traceability and lifelong learning enable learners to connect the multiple concepts they have learnt in different institutes into a meaningful chain of linked knowledge items. The resulting knowledge map could allow students to view topics in their previous learning activities (tracked through the blockchain) that are related to the current one. Students could also choose to revise on these previous topics in preparation for a new learning task. Teachers on the other hand can use this tool to access the preparedness of students, missing knowledge nodes, assign revision tasks and view outcomes as shown in figures 4.11. Administrators may also use these tools check the contents that makeup a given transcript or certificate. This way, course contents can be verified from the students'

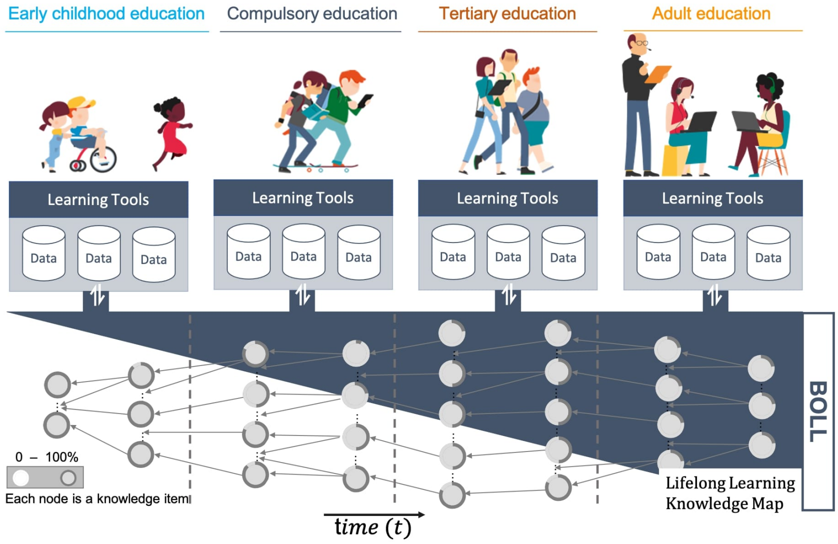


Figure 4.18: Lifelong learning traceability (Ocheja et al., 2020)

records perspective and not just the curriculum document.

4.2.8 Discussion

A. Addressing the results of the needs analysis

During the needs analysis with the teachers, it was revealed that access to past learning records is important especially for enabling personalized learning and assessing a student's ability. Our proposal meets this requirement as it connects learners' learning records across different systems. We also implemented visualization for viewing a learner's current knowledge on prerequisite concepts and a means for teachers to conduct revision and reflection activities; another need mentioned by the teachers. In order to enable access to more data, which is also a limitation mentioned by one of the teachers, the adoption of our proposed framework at the government level could help overcome data scarcity problems. However, this topic of discussion is beyond the scope of the present paper.

The teachers also mentioned that they needed a way access information such as students' engagement, learning materials used and usage frequency, and the difficulty level of quiz questions solved by the students. Our proposed visualization framework provides access to such data. While our proposal does not provide a blanket implementation to access all possible kinds of data, our proposal relies on granular data from the learning tools used by the students. This means that it is possible to extend our implementation to reflect additional information which could be retrieved from the underlying data on the blockchain. Thus, we recommend that education blockchains should also provide a way to analyze the data collected towards retrieving additional insights. We share in the views of these teachers ranging from solving the cold-start problem to enabling provision of personalized learning contents. In future work, we hope to qualitatively measure how our proposed system impacts on the performance of students whose past learning records have been connected and made accessible through the blockchain.

B. Visualization of education blockchain data

Our research reveals that most education blockchain proposals have continually focused on reporting specific credentials and in most cases only certificates, diplomas and transcripts. With such a focus on one data-point type of academic records, it is difficult to realize visualizations similar to the common Learning Analytics Dashboards (LAD's). This has a consequent effect of being unable to apply such data to multiple aspects of learning. In this work, we have presented the BOLL platform that collects and connects various types of academic records. With these kinds of data, we can provide adequate data for intelligent systems to provide better personalized learning. We recommend that research on blockchain in education should include aspects that focus on usefulness within classes for various courses and how education blockchain systems may impact learning activities.

C. Relevance of past learning records

Our interview with multiple teachers during this research has revealed that access to these kinds of data is desirable for various reasons. The teachers acknowledged the usefulness of our proposal and made additional requests such as being able to view students' engagement trends from past learning environments. This supports our initial argument that education blockchains need to provide more than just reporting and validating credentials. The usefulness of these information in supporting learning activities is also important. While we have made some proposals, there are various opportunities for innovation such as mining of education blockchain data to find trends and patterns that may improve learning outcome.

D. Adoption in various learning activities

Our perspective on education blockchain has always been on the ability to use such systems to support and improve learning outcome. However, it is often a challenge to decide how to introduce these systems into various learning activities. For instance, some of the teachers we interviewed mentioned that they have this challenge but remain curious on using such a system. To find meaningful ways to make use of education blockchain records to support learning activities, it is important to enable active collaborations between teachers and researchers. While researchers may be able to provide the tools and methods, teachers can more accurately provide information on what situations exist or are important to consider. This type of collaboration can be enabled through participatory design method when developing education blockchain tools.

E. Complexities with blockchain

One major reason for the lack of education blockchain data visualization is the poor scalability of most blockchain solutions as this has been a problem to adoption of blockchain in education (Alammary et al., 2019). With respect to writing learning logs on the blockchain, the BOLL platform also suffers from the scalability challenge of blockchains. For instance, to add a learning log to the BOLL network will take from 15 seconds to 2 minutes (Ocheja, Flanagan, Ueda, et al., 2019). The problem with this is that learning logs are generated at a much faster rate than 15 seconds. Therefore, it is necessary to determine the best approach to write these logs on the BOLL network. One way we have identified to solve this problem, is to mine only representative learning logs to the blockchain and also to batch multiple learning logs in a single transaction. Initial experiments with this approach showed significant improvement. For example, over 1 million records which would take more than a year to write to the blockchain were transferred over a two weeks period using mining of representative learning logs.

F. Education blockchains interoperability

The interoperability of education blockchains is an important consideration as we have seen in existing research (Kontzinos et al., 2020; Schmidt, 2016) that education blockchains often focus on different aspects and vary in implementation. This range from what type of academic records are being held on the blockchain to the type of blockchain used. One solution to this problem is to form learning consortium that can federate how different kinds of educational records are reported on the blockchain. In the spirit of true decentralization, the rules proposed by the consortium can be enforced on the blockchain with each member node voting to comply or not. Although this may come at the expense of being able to report certain kinds of information or not, it is important to provide standardization across education blockchains.

4.3 Investigating relevance of prior learning data connected on Boll

4.3.1 Aim and research questions

Teachers often face a common problem of not knowing the past learning engagements of their students. While final grades or scores may be contained in academic transcripts, it is difficult to measure students' engagement from transcripts. Trowler (2010) defines student engagement as the interaction between the time, effort and other relevant resources invested by students and institutes towards optimizing learning experience and to enhance students' performance. The differences in learning purposes, preferences, and motivations of students can result in different types of engagement behaviour during learning which may in-turn affect their performance (Li & Tsai, 2017). Previous research has shown that students' engagement in the learning environment is closely related to their learning outcome (M. Hu & Li, 2017; Lu et al., 2017). Thus, giving teachers access to their students past engagement could equip them with information about the possible challenges students may face, eliminate repetitive learning, how to adapt learning contents and provide support to students with prior low engagement.

To measure students' engagement at different times, it becomes necessary to access and analyze their total experience while learning at an institute. However, access to students' learning data after they change school is often difficult. This is largely attributed to the heterogeneous nature of learning systems and the lack of transferability of lifelong learning logs across schools (Baker et al., 2019). The advent of decentralized technologies such as the blockchain opens up new ways to address this problem. (Ocheja et al., 2018) proposed a blockchain of learning logs platform (BOLL) that can connect learning behaviour logs of students across different schools on a secure and immutable ledger. While the BOLL system solves the problem of learning data continuity, this paper presents a first of its kind research on providing teachers access to insights drawn from their students prior learning data such as engagement and learning outcome. In this work, we use the BOLL platform to provide teachers access to their students past learning engagements and investigate the relevance of students' past learning behaviour logs. For example, when students move from JHS 3 to HS 1, their HS 1 teacher is given access to the students' past learning behaviour logs. However, the teacher does not have data analytics skills to know if the learning behaviour logs have any effect on the final scores obtained.

Our main argument is that it is not enough to provide access to past learning logs: the relevance of such data should also be communicated to the stakeholders. This is important because in most cases, stakeholders do not have the required data analytics to carry out such investigations on their own. We also provide a first of its kind access to the learning materials and assessment data (questions, students' and teachers' solutions) used by the student at their previous school using the marketplace (Boll-M) feature of Boll (Ocheja, Flanagan, & Ogata, 2019). Specifically, this paper is focused on answering the following research questions:

1. What are the engagement levels of students at a past learning environment?

2. How relevant are these engagements to students' past learning outcome?

3. How can teachers access additional information about learning outcomes?

4.3.2 Methodology

In this research, we use the Boll system Ocheja, Flanagan, Ueda, et al. (2019) to connect the learning behaviour logs of students across two schools in Japan. We first setup the Boll system, connect it to the Learning Records Store (LRS) of the Junior High School (JHS) and assign a blockchain address to each student. The Boll system also keeps track of each student's ID at that school. This is then used to identify the records to be transferred when the student change school. When students in these schools move from the JHS 3 to High School (HS) 1 (a different school), we also transfer their past learning logs on the BookRoll system Flanagan and Ogata (2018) to their new school. The HS also has a similar setup of the Boll system with connections to the LRS. For this study, we analyzed the learning behaviour logs of 109 students in JHS 3 Mathematics course in 2020 academic year who are currently in HS 1 and have enrolled in the HS 1 Mathematics course in 2021 academic year.

Our analysis includes: engagement behaviour cohorts, temporal and spatial change in engagement and learning contents visualization. We measure engagement as a sum of different student behaviours categorized in to 5 dimensions: self-evaluation (S_e), cognitive behaviour (C_b), backtracking behaviour (B_b), time commitment (T_c) and content progress/completion (C_p). We define self-evaluation (S_e) as the students' ability to evaluate correctly their own solution to quiz questions. S_e is calculated as a fraction of the quiz answers from the student which were correct and rightly marked as correct by the student. Cognitive behaviour (C_b) is a measure of the students' cognitive action through cognitive indicators such as yellow and red markers added on learning materials through the BookRoll system Akçapınar et al. (2019). The backtracking behaviour (B_b) is an indication of how often students revisit concepts in order to improve their understanding or master such concepts. This is calculated as a weighted sum of total previous page visit actions divided by the total next page visit actions and the total previous page visit actions Yang et al. (2021). Time commitment (T_c) is a measure of how often students study and it is calculated as the weighted sum of the total time, total number of content usage events and the total number of unique days students used the contents of the course. Content progress/completion (C_p) is a measure of how students advance towards completing the study materials. It is calculated as the weighted sum of total open and next page actions and total sum of long and short events. It is important to note that the parameters of each engagement metric were percentile rank of their actual values. Thus, student overall engagement is calculated as:

$$\text{Engagement} = S_e + C_b + B_b + T_c + C_p$$

In Table 4.4, we show a summary description of the dataset for the JHS 3 Math course in 2020. The engagement metrics previously discussed were extracted from the dataset of the students who took the final exam and were graded. The engagement score was used to divide into quartile groups of 4 different engagement levels: Very High ($\geq 75th\ percentile$), High ($\geq 50th\ percentile$), $Low(\geq 25th\ percentile$) and Very Low ($<$ 25th percentile) using percentile rank. We then proceeded with ensuring the data met the assumptions of a one-way Analysis of Covariance (ANOVA) before performing conducting a test for a significant difference in the mean score for each engagement level. Finally, we developed 4 visualizations for teachers to view students' past engagement showing information such as: learner profile, group engagement, temporal and spatial engagement change and learning materials used.

Table 4.4: Description of the dataset.

	No. of Students	Total logs	No. of Students graded
Group A	40	123,678	38
Group B	40	98,080	38
Group C	40	125,619	33

4.3.3 Results

We implemented four different visualizations for stakeholders to view the past engagement of students. The learner profile shown in Figure 4.19 gives a comprehensive summary of a student's past engagement and their achievements. This can also tell the teacher if the past engagement is correlated to the student's score or not. For each of the assessments, one can also view the student's solution as well as the correct solution. The Engagement Transition in Figure 4.20 gives stakeholders ability to view change in engagement level of a group of students using iSAT (Majumdar & Iyer, 2016). For example, teachers can check transition across a period of time to know when (or at what point in the past) a student's learning behaviour changed (improved or needs intervention). The teacher can also compare engagement changes across courses, contents or activities.

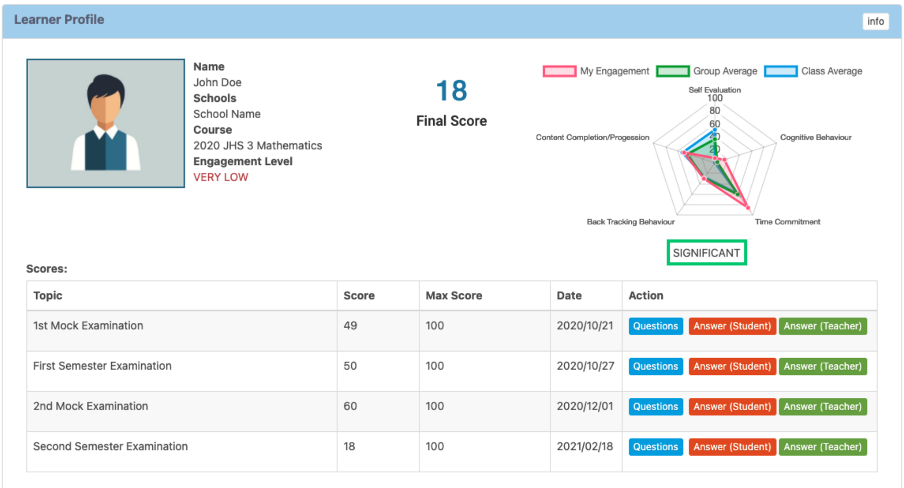


Figure 4.19: Learner profile.

The engagement groups visualization in Figure 4.21 enables stakeholders to view engagement profile of different engagement cohorts in the class and to know what characteristic are prominent among different cohorts. One can also view the details of each student

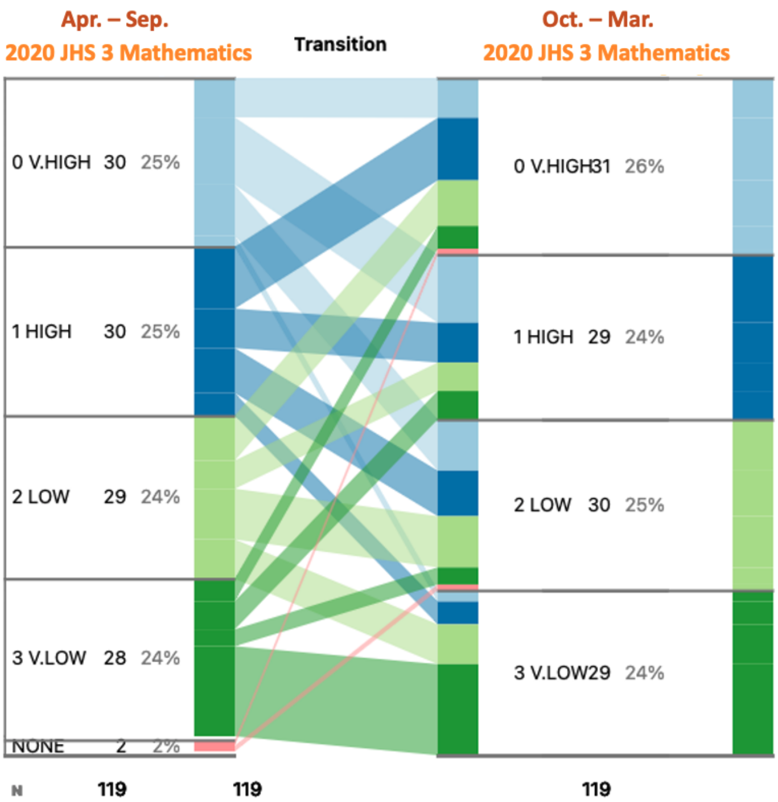


Figure 4.20: Temporal change in Engagement level.

in each cohort and assign specific tasks such as revisions and assessment retake. The learning materials interface show in Figure 4.22 provides stakeholders a way to access the learning materials students have used in the past including: textbooks, quiz questions, students' solutions and lecture slides. Figures 4.19 to 4.22 are from a real implementation of the Boll system currently deployed at a school in Japan.

Before carrying out an Analysis of Covariance (ANOVA) between the engagement levels and score, a Shapiro-Wilk test was conducted to determine the normality of the data. The result (0.99, $p > 0.05$) revealed that the score data across the different engagement levels followed a normal distribution. A further test for homogeneous variance using Levene's test indicated homogeneity of variances across the different engagement levels ($F\ (3,105) = 2.272$, $p > 0.05$). We then conducted a parametric one-way ANOVA to determine whether the mean scores of all engagement levels are different. The result ($F(3,105) = 3.783$, $p < 0.05$) indicated a significant difference in the mean scores for all engagement levels. A further post-hoc test using the Games-Howell test (due to unequal sample sizes) showed that the difference between very high and very low engagement

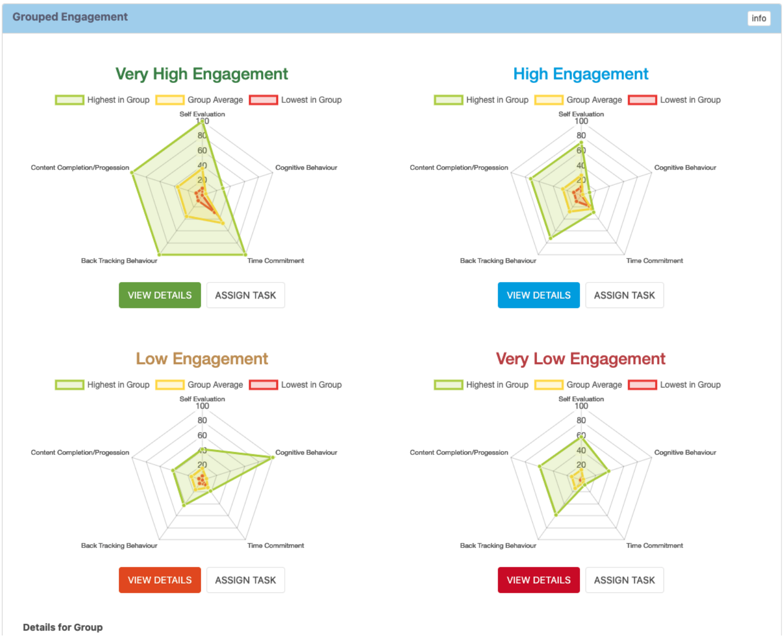
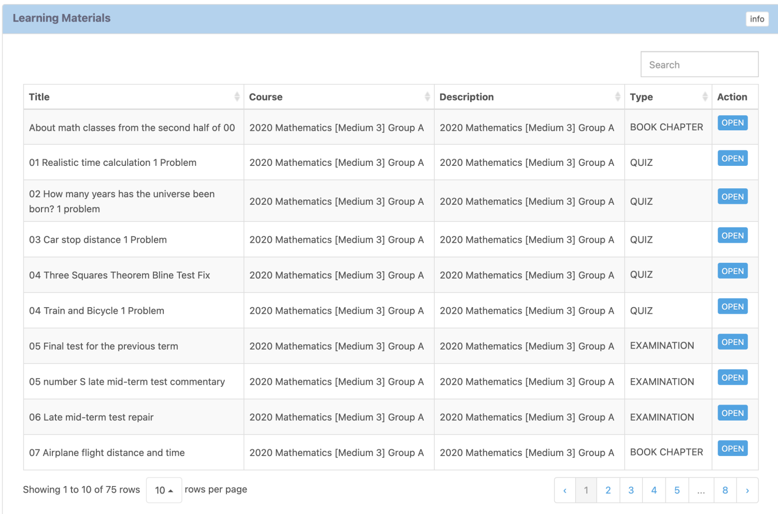


Figure 4.21: Detail profiles of Engagement Groups.



Learning Materials

Search

Title	Course	Description	Type	Action
About math classes from the second half of 00	2020 Mathematics [Medium 3] Group A	2020 Mathematics [Medium 3] Group A	BOOK CHAPTER	OPEN
01 Realistic time calculation 1 Problem	2020 Mathematics [Medium 3] Group A	2020 Mathematics [Medium 3] Group A	QUIZ	OPEN
02 How many years has the universe been born? 1 problem	2020 Mathematics [Medium 3] Group A	2020 Mathematics [Medium 3] Group A	QUIZ	OPEN
03 Car stop distance 1 Problem	2020 Mathematics [Medium 3] Group A	2020 Mathematics [Medium 3] Group A	QUIZ	OPEN
04 Three Squares Theorem Bline Test Fix	2020 Mathematics [Medium 3] Group A	2020 Mathematics [Medium 3] Group A	QUIZ	OPEN
04 Train and Bicycle 1 Problem	2020 Mathematics [Medium 3] Group A	2020 Mathematics [Medium 3] Group A	QUIZ	OPEN
05 Final test for the previous term	2020 Mathematics [Medium 3] Group A	2020 Mathematics [Medium 3] Group A	EXAMINATION	OPEN
05 number S late mid-term test commentary	2020 Mathematics [Medium 3] Group A	2020 Mathematics [Medium 3] Group A	EXAMINATION	OPEN
06 Late mid-term test repair	2020 Mathematics [Medium 3] Group A	2020 Mathematics [Medium 3] Group A	EXAMINATION	OPEN
07 Airplane flight distance and time	2020 Mathematics [Medium 3] Group A	2020 Mathematics [Medium 3] Group A	BOOK CHAPTER	OPEN

Showing 1 to 10 of 75 rows 10 rows per page

‹ 1 2 3 4 5 ... 8 ›

Figure 4.22: Past learning materials transfer across schools.

levels is significant ($p < 0.05$) as presented in Table 4.5. The implication of this result is that very low and very high engagement levels are indicative of the final performance of students and provide actionable insights for guiding future teaching and learning.

4.3.4 Discussion

This work makes an important contribution of investigating and informing stakeholders the effect of students' prior engagement on their final scores at a different learning environ-

Table 4.5: Post-hoc test (Games-Howell) results of scores between engagement levels (mean difference, standard error)

	N	Score(μ)	SD	Very High	High	Low	Very Low
Very High	28	59.64	16.01	-	3.72 (4.12)	6.50 (4.02)	14.53 (5.06)*
High	27	55.93	14.51		-	2.78 (3.85)	10.82 (4.92)
Low	27	53.15	13.76			-	8.04 (4.84)
Very Low	27	45.11	21.07				-

ment. Such information makes it possible for teachers to provide specific interventions at the start of a new class without having to wait to collect some data in the first few weeks. Although the results from our analysis only revealed a significant correlation between the scores and engagement of very high and very low engagement students, we propose this type of analysis to be performed when providing stakeholders with learning logs from a different learning environment.

In addition to engagement and final scores, this work provided access to resources such as the students' solution to examination questions and learning materials used. Access to this type of data give teachers additional information about the students' ability, and challenges with respect to the assessment questions. We acknowledge that in some cases, other contextual information may be required to correctly interpret the engagement measures extracted from the learning logs. Also, students may have received other scores different from the final score. It may be useful to consider how the students' engagement at intervals preceding other assessment affected their performance.

4.4 Supporting students' higher education enrolment on Boll

4.4.1 Aim and research questions

4.4.2 Methodology

To evaluate the usefulness of the proposed framework, we use the assessment results of students in a High School in Japan to build a decentralized model that can support students in making data-informed enrolment decisions. We begin by connecting the learning infrastructure at the selected school to BOLL (Ocheja, Flanagan, Ueda, et al., 2019). We design the smart contract for this LAS to require the score data of high school students from various tests conducted in High Schools 1 to 3 in Mathematics, English, Japanese, and National subjects to be provided as inputs to the model. For these set of input, the compute function will return the top n program categories a student's cognitive ability is most suitable and how the student compare to their peers in the different program categories. Also, the smart contract has another compute function that returns the performance distribution across each program category and the mean score required for successful entry.

Table 4.6: Dataset for training initial model (2010 - 2020).

Program Type	No. Students	Math		English		Japanese		National	
		μ	σ	μ	σ	μ	σ	μ	σ
I Education...	1116	68.17	12.57	57.27	14.76	55.86	14.31	59.30	13.02
II Science...	169	52.17	16.15	48.29	16.22	51.58	15.56	47.89	14.23
III Sports...	164	62.15	14.40	54.88	15.46	55.85	15.20	55.64	14.36
IV Medicine...	870	60.13	14.00	56.63	15.20	58.26	14.29	56.60	13.88
V Economics...	559	65.63	19.96	61.75	18.69	59.34	18.35	60.85	18.42
VI Others	173	57.69	14.11	58.02	15.70	59.22	15.26	56.39	14.07

We use the assessment results of past students (2010 - 2020) shown in Table 4.6 and the program category they successfully enrolled in to build the initial model using Random Forest classifier that can predict enrolment likelihood and recommend the most suitable options. There are six program categories: I - *Education, Literature/Humanities, Foreign Languages*; II - *Science, Engineering, Agriculture*; III - *Sports, Arts, Arts & Sciences*; IV - *Medicine, Pharmacy, Dentistry and Nursing*; V - *Economics, Law*; and VI - *Others*. To reduce the data imbalance across the 6 categories we used SMOTE Chawla et al., 2002 to generate synthetic samples with a fair balance between minority and majority samples.

4.4.3 Results

The Area Under the Curve (AUC) scores presented in Table 4.7 showed that the resulting models performed 70% and above for the cases in bold. Further analysis of these cases revealed that the model makes a better judgement where comparing between Art vs Science oriented program types. For example, the classifier for program type I (Arts) vs program type II (Science) had an AUC score of 82%. But does it not state what we already know: Arts and Science oriented programs are different? Yes, it does but surprisingly, the data reveals that even though both program types are different, over 200 students still went ahead to apply to both. This comes with the consequence of students spreading themselves too thin and finding it difficult to make smart choices and preparations. To support students to plan better, we also provide a visualization in Figure 4.23 as an example of the visual layer of our proposed framework. Through this interface, students can see their current performance, compare to previous students who have enrolled in certain program types and what additional grounds they need to cover in each subject in order to have a better chance at enrolling.

Table 4.7: AUC for predicting enrolment decision based on score data.

Program Type	I	II	III	IV	V	VI
I	1					
II	**0.82**	1				
III	0.65	**0.78**	1			
IV	**0.70**	0.60	0.62	1		
V	0.56	**0.81**	**0.70**	**0.72**	1	
VI	**0.80**	0.69	**0.74**	**0.70**	**0.80**	1

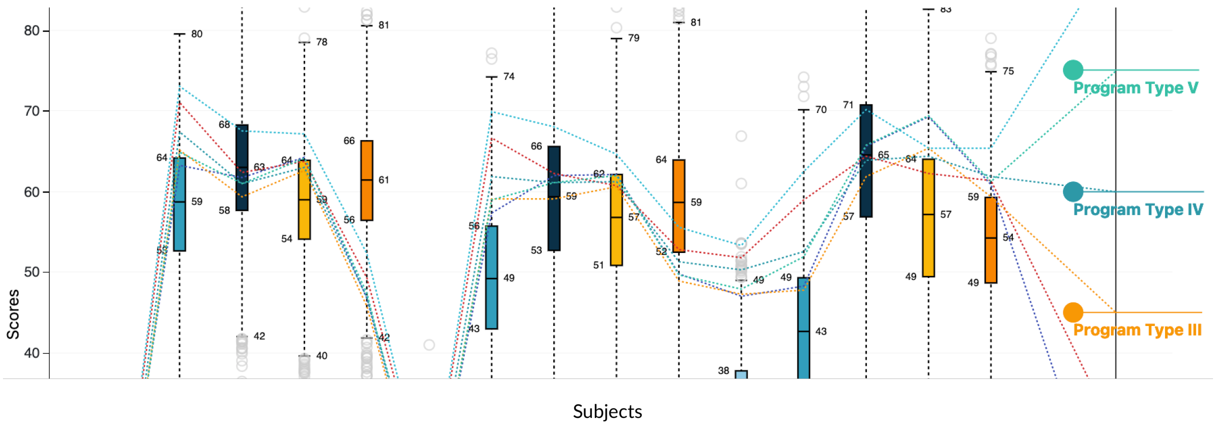


Figure 4.23: Interactive visualization to support enrolment decision and preparation

4.4.4 Discussion

In this work we conceptualized a framework for managing assessment results on a decentralized network using the blockchain. We proposed three (3) main layers: service, data and visual. These layers can then interface with the blockchain network to provide distributed services, leverage on decentralized data and provide insights to support teaching and learning. Also, we presented results from an initial use-case which suggest that valuable insights can be retrieved from assessment results of past and current learners when available. There is also a lot of potential to expand beyond education in schools as the BOLL could cross into private sector employment and training. The advantage of a system like BOLL to realize this is the underlying decentralized architecture that allow different parties to retain existing infrastructure and only connect their data source to BOLL as shown in the architecture proposed in this paper.

Chapter 5

Conclusion and future work

To conclude, we will first present the main findings and contribution of this research and end by proposing directions for future research.

5.1 Findings

In this thesis, we identified the existing and perturbing cold-start problem faced by learning analytics systems due to the lack of learning data-continuity. To solve this problem, we proposed, designed and implement the Boll system to enable transfer of learning records and digital contents across school. The Boll system also provides features to support privacy, learning traceability, verification, immutability and consistency of reported data on the blockchain. We further implemented visualizations and interfaces to enable use of distributed academic records to support teaching and learning beyond a single institution. Finally, we explored the relevance of the proposed system by conducting experiments with key stakeholders and evaluating the relevance of connected learning logs to lifelong learning. The main findings related to the research questions are presented below.

5.1.1 Findings from design

RQ1.1. How to connect distributed lifelong learning logs of students across different schools?

This research revealed that it is possible to connect lifelong learning logs of students across different schools by using the blockchain. We showed that on the blockchain, each block can represent a learner's learning activity and the blocks can be chained in a manner that the most recent block points to the preceding block in a traceable way. Also, this work proposed smart contracts that contains rules on how different schools can join the education blockchain and register their students. We also designed smart contracts that define how learning data is reported and how various stakeholders can protect and transfer their data.

RQ1.2. How to enable the transfer of digital learning materials across different schools with intellectual property protection and transparent use?

Our work demonstrated a framework that can enable learners transfer their learning contents when they change and at the same time, protect the intellectual property of authors and publishers using smart contracts on the blockchain. We also presented typical templates for drafting various kinds of policies that can foster such decentralized trust among authors, sponsors and users of digital contents. Our design also revealed how such solution can be connected to existing learning infrastructure that connects lifelong learning logs and engenders data continuity.

RQ1.3. What mechanisms can we use to enable exchange of information and learning analytics across schools?

To enable exchange of information and learning analytics across schools we can setup a decentralized network where nodes are approved learning institutions or relevant institutes that can support learners' goals such as content providers, authors and publishers. To enable sharing of learning analytics, we proposed a framework that allows a decentralized analysis of data connected through the blockchain to support specific objectives without comprising the privacy of concerned stakeholders.

5.1.2 Findings from impact evaluation

RQ2.1. What cold-start problems does connecting lifelong learning logs solve?

This work revealed that when we enable learning data continuity, it is possible to solve cold-start problems associated with on-boarding new learners. For example, An analysis of connected lifelong learning logs in this thesis showed that students' prior engagement levels were indicative of their final score. This directly solves the cold-start problem of not knowing what student's prior engagement levels were. Such information could be useful to recommender systems to recommend the right learning plan or engagement at the new school. Other cold-start problems include: awareness of prior learning contents used by learners, behaviour indicators and cohort membership and distribution.

RQ2.2. What are the perceptions of teachers about prior learning data, lack of and how do they use such information?

Our work also reveals that teachers consider access to prior learning data to be important but not possible before our proposal. The teachers indicated that information such as students' prior engagements, learning contents used, difficulty level of assessment problems and students' solutions also consist important pieces of information they would like to see when working with new students. They noted that such information could help them in difficult situations to diagnose correctly and provide the appropriate intervention for their students. Also, some of the teachers highlighted the uncertainties that comes with such amount of information and how it is used. Thus, care should be taken when providing prior learning data so as to avoid wrong interpretations. We provided examples of visualizations that could guide proper use and continuous stakeholder awareness.

RQ2.3. What is the relevance of distributed lifelong learning data and analytics to learners' future goals?

Distributed lifelong learning data and analytics is relevant to learners' future goals where it provide various insights on what personalized paths a student can take towards achieving set objectives. This worked presented a typical example of guiding the higher education enrolment decision of high school students using distributed score data connected through the blockchain. Our analysis revealed that without such distributed analysis, learners run at risk of applying to many programs that do not match their competence, spreading themselves too thin and potentially failing to get into the right program. We found that by using the lifelong learning data connected by our proposals in this thesis, we can help students prepare better, make smarter decisions and improve their chance of getting into programs that at very least, match their competence.

5.2 Implications

The findings of this work have implications for education and learning analytics research. First, by solving the cold-start problem, this thesis provides access to a data-rich environment that can reveal meaningful insights useful for supporting learners from the onset when they change school. Also, learning analytics tools and processes such as those found in recommender and intervention systems can use such data to improve the timeliness and accuracy of the personalizations offered in various learning scenarios. Second, the contributions of this thesis on traceability and transfer of learning logs and digital contents, opens a new perspective to how learning is designed, shared and revised. Unlike before, authors, learning content designers, and learners can now share useful information such as the quality of contents, suitability, and useful or problematic aspects that require attention or revision. Third, the decentralized learning analytics framework proposed in this work, offers new opportunities for research and production systems to co-exist, access similar datasets and provide timely feedback to learners in a decentralized and privacy-preserving manner. Thus, eliminating the existing challenge of limited datasets and privacy restrictions when tackling multisource-dependent learning analytics tasks.

The results from this thesis also have further implications for research on enabling lifelong learning. While this thesis may have focused on lifelong learning from the perspective of education systems and institutions, the ubiquitous nature of learning means that infinite number of sources exist. For example, informal learning such as in social contexts need to be accounted for and included in lifelong learning. Through the proposed design of the Boll system, this thesis sets a fundamental template that can be used to drive future integration of other learning sources: both formal and informal.

The technical contributions in this thesis present some key implications for the design of future blockchains and blockchain-based systems. Our experiments on the use of blockchain in education context and the results implies the need to consider use-case scenarios such as this in the design of future blockchains. For example, the concept of gas fees and double spending may be suitable for financial scenarios but hardly is this directly relevant in the writing and transfer of learning logs and contents. Therefore, domain specific constraints should be considered in future implementations of general or specific purpose blockchains. Also, the contributions of this thesis in adapting a general purpose blockchain to a specific use-case sets the path for other domains to understand the intricacies and requirements when implementing such a system. Examples of these contributions include our design of smart contracts, integrations with existing systems, visualizations of blockchain data and various use-cases through decentralized analytics.

The findings of this work have ripple effects on other domains such as healthcare, knowledge management, AI ethics, design of intelligent and agent-based systems where similar settings of user data continuity, the need for privacy, control, data transfer and analytics are desirable. For instance, the implications of this work for healthcare research spans through data privacy to personalized medicine. The similarity between patients mobility across hospitals and students frequent change of school speaks to the applicability of our work in healthcare. Another similar aspect is personalization. Healthcare research is gradually moving towards provision of personalized medicine where each patient is given a unique treatment specific to their health conditions and needs different from the current approach of one medication for same health condition. Our work on learning personalization using distributed learner data on the blockchain could inspire new efforts that can help in realizing the data protection, continuity and analytics aspects of personalized medicine.

5.3 Limitations

Although the importance of connecting and transferring lifelong learning logs and contents have been demonstrated, some limitations should be noted. First, a far-reaching implementation and adoption of the proposed Boll system is lacking. It is necessary to advance the adoption of the proposed system as such decentralized systems thrive largely based on adoption. Second, the results reported in this thesis involved participants in K-12 education and partly higher education. This work is yet to be considered across a full spectrum of elementary school to postgraduate education or other learning environments. Finally, decentralized learning analytics is a term first proposed in this thesis. Its use, relevance and impact needs to be further evaluated. Such evaluations will be helpful in revealing additional ways to collectively support learners to achieve their learning goals.

The computational complexity of blockchain operations remains a bottleneck. For example, current blockchain implementations are known to severely suffer from scalability, and high latency and throughput of transactions are rarely guaranteed. These problems have not been considered in this thesis as they are beyond the scope of the current work. However, solving these limitations will be crucial to the adoption, scalability and availability of education blockchains.

5.4 Conclusion

This thesis conducted a theoretical and practical investigation on connecting distributed learning data and analytics with studies on needs, designs and evaluations. The findings suggest that connecting learning data of learners across different schools can be beneficial to both teaching and learning, solves the cold-start problem and further enables lifelong learning and analytics. This research also provides concrete support for enabling personalized learning at scale and also enables cross-border analytics of lifelong learning. Hence, this work can create a paradigm shift in data-driven education to a decentralized approach where all institutes can work collectively to impact knowledge on the learner.

5.5 Future work

In future work, greater focus on standardized formats for representing permissions on the blockchain is necessary. The scalability of the current Boll system should also be continuously revised as research on blockchain advances. This will help to ensure that the Boll system can handle being implemented as a wide reaching system. While we acknowledge that the time taken to write learning records to the blockchain currently is not suitable for real-time access-based systems, we recommend its usage in scenarios where transition from one institution to another occurs over a given period of time that is within the waiting time as earlier presented.

Also, we recommend a continuous effort in improving the awareness and use of digital tools by teachers. To foster the adoption of the Boll system by teachers, it is necessary to demonstrate to teachers how the Boll system can solve certain day-to-day challenges faced by teachers such as; probing learning difficulty and understanding the individual needs of each student or the whole class as a group from prior data. Furthermore, it is necessary to evaluate the use of decentralized learning analytics and its visualizations with key stakeholders: focusing on conducting in-class experiments and usability testing with actual use of our proposed visualizations to support teachers and students' learning activities. This will help to reveal other affordances of decentralized learning analytics and how to better support lifelong learning.

Publications

Journals (peer-reviewed)

1. Patrick Ocheja, Brendan Flanagan, Hiroshi Ueda, & Hiroaki Ogata. (2019). Managing lifelong learning records through blockchain, Research and Practice in Technology Enhanced Learning 14 (1), 4.

2. Patrick Ocheja, Brendan Flanagan, Hiroaki Ogata & Solomon S. Oyelere. (2022). Visualization of education blockchain data: trends and challenges. Interactive Learning Environments, 1-25.

3. Patrick Ocheja, Friday J. Agbo, Solomon S. Oyelere, Brendan Flanagan, & Hiroaki Ogata. Blockchain in Education: Systematic Review and Practical Case Studies. IEEE Access, (conditionally accepted).

4. Patrick Ocheja, Brendan Flanagan, & Hiroaki Ogata. A BigQuery Interface for Decentralized Education Data on the Blockchain. Future Generation Computer Systems, (to be submitted in July 2022).

International conferences (peer-reviewed)

1. Patrick Ocheja, Brendan Flanagan, & Hiroaki Ogata. (2018). Connecting Decentralized Learning Records: a Blockchain Based Learning Analytics Platform. In Proceedings of the 8th International Conference on Learning Analytics and Knowledge, pp. 265-269. Sydney, Australia.

2. Patrick Ocheja, Brendan Flanagan, & Hiroaki Ogata. (2019). Decentralized E-Learning Marketplace: Managing Authorship and Tracking Access to Learning Materials Using Blockchain. The 2nd International Cognitive Cities Conference, Springer CCIS. pp. 526-535. Kyoto, Japan.

3. Patrick Ocheja, Brendan Flanagan, Louis Lecailliez, & Hiroaki Ogata. (2020). Improving Learning Analytics and Student Performance through Connected Lifelong Learning on the Blockchain. In Companion Proceedings of the 10th International Conference on Learning Analytics and Knowledge, (DC@LAK20), pp. 608-616.

4. Patrick Ocheja, Brendan Flanagan, Solomon S. Oyelere, Louis Lecailliez, & Hiroaki Ogata. (2020). A Prototype Framework for a Distributed Lifelong Learner Model. The 28th International Conference on Computers in Education, pp. 261-266.

5. Patrick Ocheja, Brendan Flanagan, & Hiroaki Ogata. (2021). Blockchain in Education: Connecting Learning Records and Contents through the Blockchain. In Proceedings of Blockchain in Kyoto 2021 (BCK21) (p. 011006). JPS Conference Proceedings.

6. Patrick Ocheja, Brendan Flanagan, Rwitajit Majumdar, & Hiroaki Ogata. (2021). Investigating Relevance of Prior Learning Data Connected through the Blockchain. In Proceedings of the 27th International Conference on Computers in Education (ICCE2021), pp. 279-284.

7. Patrick Ocheja, Brendan Flanagan, & Hiroaki Ogata. Assessment Results on the Blockchain: A Conceptual Framework. In International conference on artificial intelligence in education (in-press).

References

Abelson, H. (2008). The creation of opencourseware at mit. *Journal of Science Education and Technology*, *17*(2), 164–174.

Akçapınar, G., Hasnine, M. N., Majumdar, R., Flanagan, B., & Ogata, H. (2019). Exploring the relationships between students' engagement and academic performance in the digital textbook system. *Proceedings of the 27th International Conference on Computers in Education (ICCE2019)*, 318–323.

Alammary, A., Alhazmi, S., Almasri, M., & Gillani, S. (2019). Blockchain-based applications in education: A systematic review. *Applied Sciences*, *9*(12), 2400.

Anderson, A., Nadalin, A., Parducci, B., Engovatov, D., Lockhart, H., Kudo, M., Humenn, P., Godik, S., Anderson, S., Crocker, S., et al. (2003). Extensible access control markup language (xacml) version 1.0. *OASIS*.

Apereo. (2016). *Openlrw: Open learning record warehouse*. Retrieved April 8, 2018, from https://github.com/Apereo-Learning-Analytics-Initiative/OpenLRW

Arnold, K. E., Lonn, S., & Pistilli, M. D. (2014). An exercise in institutional reflection: The learning analytics readiness instrument (lari). *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, 163–167.

Azaria, A., Ekblaw, A., Vieira, T., & Lippman, A. (2016). Medrec: Using blockchain for medical data access and permission management. *Open and Big Data (OBD), International Conference on*, 25–30.

Baker, R. S. et al. (2019). Challenges for the future of educational data mining: The baker learning analytics prizes. *Journal of Educational Data Mining*, *11*(1), 1–17.

Bakharia, A., Kitto, K., Pardo, A., Gašević, D., & Dawson, S. (2016). Recipe for success: Lessons learnt from using xapi within the connected learning analytics toolkit. *Proceedings of the sixth international conference on learning analytics & knowledge*, 378–382.

Barnes, T., & Stamper, J. (2008). Toward automatic hint generation for logic proof tutoring using historical student data. *International Conference on Intelligent Tutoring Systems*, 373–382.

Beatty, B., & Ulasewicz, C. (2006). Faculty perspectives on moving from blackboard to the moodle learning management system. *TechTrends*, *50*(4), 36–45.

Bell, K. M., Bleau, D. N., & Davey, J. T. (2011). Push notification service [US Patent 8,064,896].

Bienkowski, M., Brecht, J., & Klo, J. (2012). The learning registry: Building a foundation for learning resource analytics. *Proceedings of the 2nd international conference on learning analytics and knowledge*, 208–211.

Bonifati, A., Chrysanthis, P. K., Ouksel, A. M., & Sattler, K.-U. (2008). Distributed databases and peer-to-peer databases: Past and present. *ACM SIGMOD Record*, *37*(1), 5–11.

Bracamonte, V., & Okada, H. (2017). A review of blockchain technology applications for academic institutions. *IEICE Technical Report for Technical Committee on Social Implications of Technology and Information Ethics (SITE)*, *117*(340), 11–14.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77–101.

Cantor, S., & Scavo, T. (2005). Shibboleth architecture. *Protocols and Profiles*, *10*, 16.

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, *4*(5-6), 318–331.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, 321–357.

Chen, G., Xu, B., Lu, M., & Chen, N.-S. (2018). Exploring blockchain technology and its potential applications for education. *Smart Learning Environments*, *5*(1), 1.

Chuang, J. C.-I., & Sirbu, M. A. (2000). *Network delivery of information goods: Optimal pricing of articles and subscriptions*. Proc. Internet Publishing Beyond: Econom. Digital Inform. Intellectual Property. Kennedy School of Government, Harvard University Cambridge, MA.

ConsenSys. (n.d.). Ethereum network status. https://ethstats.net/

Consortium, I. G. L. (2015). Caliper analytics. http://www.imsglobal.org/activity/caliper

Consortium, I. G. L. (2013). *Learning measurement for analytics whitepaper*. Retrieved April 7, 2018, from https://www.imsglobal.org/sites/default/files/caliper/IMSLearningAnalyticsWP.pdf

Consortium, I. G. L. (2017). Comprehensive learner record. Retrieved May 28, 2018, from https://www.imsglobal.org/activity/comprehensive-learner-record

Council, N. R. et al. (2000). *The digital dilemma: Intellectual property in the information age*. National Academies Press.

Dawson, S., Poquet, O., Colvin, C., Rogers, T., Pardo, A., & Gasevic, D. (2018). Rethinking learning analytics adoption through complexity leadership theory. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, 236–244.

De Angelis, S., Aniello, L., Baldoni, R., Lombardi, F., Margheri, A., & Sassone, V. (2018). Pbft vs proof-of-authority: Applying the cap theorem to permissioned blockchain.

Drachsler, H., & Greller, W. (2012). The pulse of learning analytics understandings and expectations from the stakeholders. *Proceedings of the 2nd international conference on learning analytics and knowledge*, 120–129.

Education, S. G. (2017). *Sony develops system for authentication, sharing, and rights management blockchain technology*. Retrieved September 28, 2017, from https://www.sony.net/SonyInfo/News/Press/201708/17-071E/index.html

Ehlers, U.-D. et al. (2018). Higher creduation–degree or education? the rise of microcredentials and its consequences for the university of the future. *European Distance and E-Learning Network (EDEN) Conference Proceedings*, (1), 456–465.

Emanuel, E. J. (2013). Online education: Moocs taken by educated few. *Nature*, *503*(7476), 342.

Ethereum. (2013a). *Ethereum in Go language*. Retrieved April 4, 2018, from https://github.com/ethereum/go-ethereum

Ethereum. (2013b). *The solidity contract-oriented programming language*. Retrieved July 27, 2018, from https://github.com/ethereum/solidity

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5/6), 304–317.

Flanagan, B., & Ogata, H. (2017a). Integration of learning analytics research and production systems while protecting privacy. In W. Chen (Ed.), *Proceedings of the 25th international conference on computers in education. New Zealand: Asia pacific society for computers in education* (pp. 333–338).

Flanagan, B., Majumdar, R., Akçapınar, G., Wang, J., & Ogata, H. (2019). Knowledge map creation for modeling learning behaviors in digital learning environments.

Flanagan, B., & Ogata, H. (2017b). Integration of learning analytics research and production systems while protecting privacy. *The 25th International Conference on Computers in Education, Christchurch, New Zealand*, 333–338.

Flanagan, B., & Ogata, H. (2018). Learning analytics platform in higher education in japan. *Knowledge Management & E-Learning: An International Journal*, *10*(4), 469–484.

Foroughi, A., Albin, M., & Gillard, S. (2002). Digital rights management: A delicate balance between protection and accessibility. *Journal of information science*, *28*(5), 389–395.

Fournier, H., Kop, R., & Sitlia, H. (2011). The value of learning analytics to networked learning on a personal learning environment. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, 104–109.

Graf, S., Ives, C., Lockyer, L., Hobson, P., & Clow, D. (2012). Building a data governance model for learning analytics. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 21–22.

Grech, A., & Camilleri, A. F. (2017). Blockchain in education.

Hickey, D., Jovanovic, J., Lonn, S., & Willis, J. E. (2015). 2nd int'l workshop on open badges in education (obie 2015): From learning evidence to learning analytics. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, 392–393.

Hoffman, M. R., Ibáñez, L.-D., Fryer, H., & Simperl, E. (2018). Smart papers: Dynamic publications on the blockchain. *European Semantic Web Conference*, 304–318.

Hu, M., & Li, H. (2017). Student engagement in online learning: A review. *2017 International Symposium on Educational Technology (ISET)*, 39–43.

Hu, V. C., Ferraiolo, D., Kuhn, R., Friedman, A. R., Lang, A. J., Cogdell, M. M., Schnitzer, A., Sandlin, K., Miller, R., Scarfone, K., et al. (2013). Guide to attribute based access control (abac) definition and considerations (draft). *NIST special publication*, *800*(162), 1–54.

36, I. J. 1. (2016). *Information technology for learning, education and training – Learning analytics interoperability – Part 1: Reference model* (Standard). International Organization for Standardization. Geneva, CH.

Janowicz, K., Regalia, B., Hitzler, P., Mai, G., Delbecque, S., Fröhlich, M., Martinent, P., & Lazarus, T. (2018). On the prospects of blockchain and distributed ledger technologies for open science and academic publishing. *Semantic web*, *9*(5), 545–555.

Jiménez-Gómez, M. Á., Luna, J. M., Romero, C., & Ventura, S. (2015). Discovering clues to avoid middle school failure at early stages. *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge*, 300–304.

Kanuka, H. (2008). Understanding e-learning technologies-in-practice. *The theory and practice of online learning*, 91.

Kennedy, G., Coffrin, C., De Barba, P., & Corrin, L. (2015). Predicting success: How learners' prior knowledge, skills and activities predict mooc performance. *Proceedings of the fifth international conference on learning analytics and knowledge*, 136–140.

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering.

Kitto, K., Cross, S., Waters, Z., & Lupton, M. (2015). Learning analytics beyond the lms: The connected learning analytics toolkit. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, 11–15.

KMi, D. D. .-. (2020). Researching the potential of blockchains. https://blockchain.open.ac.uk/experiments/

Kontzinos, C., Kokkinakos, P., Skalidakis, S., Markaki, O., Karakolis, V., & Psarras, J. (2020). Decentralised qualifications' verification and management for learner empowerment, education reengineering and public sector transformation: The qualichain project. *Mobile, Hybrid, and On-line Learning (eLmL 2020)*, 51.

Kump, B., Seifert, C., Beham, G., Lindstaedt, S. N., & Ley, T. (2012). Seeing what the system thinks you know: Visualizing evidence in an open learner model. *Proceedings of the 2nd international conference on learning analytics and knowledge*, 153–157.

Learning, A. D. (2016a). Adlnet/xapi-spec: The xapi specification describes communication about learner activity and experiences between technologies. http://github.com/adlnet/xAPI-Spec/

Learning, A. D. (2016b). *Experience api (xapi) specification*. Retrieved May 18, 2018, from http://github.com/adlnet/xAPI-Spec/

Li, L.-Y., & Tsai, C.-C. (2017). Accessing online learning material: Quantitative behavior patterns and their effects on motivation and learning performance. *Computers & Education*, *114*, 286–297.

Liu, Q., Safavi-Naini, R., & Sheppard, N. P. (2003). Digital rights management for content distribution. *Proceedings of the Australasian information security workshop conference on ACSW frontiers 2003-Volume 21*, 49–58.

Longhurst, R. (2003). Semi-structured interviews and focus groups. *Key methods in geography*, *3*(2), 143–156.

Lorch, M., Proctor, S., Lepro, R., Kafura, D., & Shah, S. (2003). First experiences using xacml for access control in distributed systems. *Proceedings of the 2003 ACM workshop on XML security*, 25–37.

Lu, O. H., Huang, J. C., Huang, A. Y., & Yang, S. J. (2017). Applying learning analytics for improving students engagement and learning outcomes in an moocs enabled collaborative programming course. *Interactive Learning Environments*, *25*(2), 220–234.

Macfadyen, L. P., Dawson, S., Pardo, A., & Gaševic, D. (2014). Embracing big data in complex educational systems: The learning analytics imperative and the policy challenge. *Research & Practice in Assessment*, *9*, 17–28.

Majumdar, R., & Iyer, S. (2016). Isat: A visual learning analytics tool for instructors. *Research and practice in technology enhanced learning*, *11*(1), 1–22.

Mandran, N., Ortega, M., Luengo, V., & Bouhineau, D. (2015). Dop8: Merging both data and analysis operators life cycles for technology enhanced learning. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, 213–217.

Mangaroska, K., Vesin, B., & Giannakos, M. (2019). Cross-platform analytics: A step towards personalization and adaptation in education. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 71–75.

May, C. (2006). *World intellectual property organization (wipo): Resurgence and the development agenda*. Routledge.

Mikroyannidis, A. (2020). Blockchain applications in education: A case study in lifelong learning.

Misiejuk, K., & Wasson, B. (2017). State of the field report on learning analytics.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The prisma statement. *PLoS medicine*, *6*(7), e1000097.

Moodle, H. (2001). Moodle learning management system. Retrieved May 28, 2018, from https://moodle.org/

Mouri, K., & Ogata, H. (2015). Ubiquitous learning analytics in the real-world language learning. *Smart Learning Environments*, *2*(1), 1–18.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. Retrieved March 18, 2018, from https://bitcoin.org/bitcoin.pdf

Ocheja, P., Flanagan, B., & Ogata, H. (2018). Connecting decentralized learning records: A blockchain based learning analytics platform. *Proceedings of the 8th international conference on learning analytics and knowledge*, 265–269.

Ocheja, P., Flanagan, B., & Ogata, H. (2019). Decentralized e-learning marketplace: Managing authorship and tracking access to learning materials using blockchain. *International Cognitive Cities Conference*, 526–535.

Ocheja, P., Flanagan, B., Ogata, H., & Oyelere, S. S. (2022). Visualization of education blockchain data: Trends and challenges. *Interactive Learning Environments*, 1–25.

Ocheja, P., Flanagan, B., Oyelere, S. S., Lecailliez, L., & Ogata, H. (2020). A prototype framework for a distributed lifelong learner model. *28th International Conference on Computers in Education Conference Proceedings*, *1*, 261–266.

Ocheja, P., Flanagan, B., Ueda, H., & Ogata, H. (2019). Managing lifelong learning records through blockchain. *Research and Practice in Technology Enhanced Learning*, *14*(1), 1–19.

Ogata, H., Li, M., Hou, B., Uosaki, N., El-Bishouty, M. M., & Yano, Y. (2011). Scroll: Supporting to share and reuse ubiquitous learning log in the context of language learning. *Research & Practice in Technology Enhanced Learning*, *6*(2), 69–82.

Okada, M., & Tada, M. (2014). Formative assessment method of real-world learning by integrating heterogeneous elements of behavior, knowledge, and the environment. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, 1–10.

Okubo, F., Yamashita, T., Shimada, A., & Ogata, H. (2017). A neural network approach for students' performance prediction. *LAK*, 598–599.

Oster, M., Lonn, S., Pistilli, M. D., & Brown, M. G. (2016). The learning analytics readiness instrument. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 173–182.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, *45*(3), 438–450.

Patton, M. Q. et al. (1980). Qualitative evaluation methods.

Piety, P. J., Hickey, D. T., & Bishop, M. (2014). Educational data sciences: Framing emergent practices for analytics of learning, organizations, and systems. *Proceedings of the fourth international conference on learning analytics and knowledge*, 193–202.

Prinsloo, P., & Slade, S. (2015). Student privacy self-management: Implications for learning analytics. *Proceedings of the fifth international conference on learning analytics and knowledge*, 83–92.

University of Nicosia. (2014). *University of nicosia: Academic certificates on the blockchain*. Retrieved May 7, 2018, from http://digitalcurrency.unic.ac.cy/certificates

The Mozilla Foundation, & in collaboration with The MacArthur Foundation, P. 2. P. U. (2012). Open badges working paper. Retrieved May 5, 2018, from https://wiki.mozilla.org/images/5/59/OpenBadges-Working-Paper%5C_012312.pdf

Rienties, B., & Toetenel, L. (2016). The impact of 151 learning designs on student satisfaction and performance: Social learning (analytics) matters. *Proceedings of the sixth international conference on learning analytics & knowledge*, 339–343.

Rubel, A., & Jones, K. M. (2016). Student privacy in learning analytics: An information ethics perspective. *The Information Society*, *32*(2), 143–159.

Samuelsen, J., Chen, W., & Wasson, B. (2019). Integrating multiple data sources for learning analytics-review of literature. *Research and Practice in Technology Enhanced Learning*, *14*(1), 11.

Schmidt, P. (2016). Blockcerts—an open infrastructure for academic credentials on the blockchain. *MLLearning (24/10/2016)*. Retrieved May 22, 2018, from https://medium.com/mit-media-lab/blockcerts-an-open%5C%5C-infrastructure-for-academic-credentials-on-the-blockchain-899a6b880b2f

Sclater, N., Peasgood, A., & Mullan, J. (2016). Learning analytics in higher education. *London: Jisc. Accessed February*, *8*, 2017.

Sharples, M., & Domingue, J. (2016). The blockchain and kudos: A distributed system for educational record, reputation and reward. In K. Verbert, M. Sharples, & T. Klobučar (Eds.), *Adaptive and adaptable learning* (pp. 490–496). Springer International Publishing.

Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S. B., Ferguson, R., Duval, E., Verbert, K., Baker, R., et al. (2011). *Open learning analytics: An integrated & modularized platform* (Doctoral dissertation). Open University Press Doctoral dissertation.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, *46*(5), 30.

Slade, S., Prinsloo, P., & Khalil, M. (2019). Learning analytics at the intersections of student trust, disclosure and benefit. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 235–244.

Stinebrickner, R., & Stinebrickner, T. R. (2014). A major in science? initial beliefs and final outcomes for college major and dropout. *Review of Economic Studies*, *81*(1), 426–472.

Suthers, D., & Rosen, D. (2011). A unified framework for multi-level analysis of distributed learning. *Proceedings of the 1st international conference on learning analytics and knowledge*, 64–74.

Szabo, N. (1997). Formalizing and securing relationships on public networks. *First Monday*, *2*(9).

Tene, O., & Polonetsky, J. (2012). Big data for all: Privacy and user control in the age of analytics. *Nw. J. Tech. & Intell. Prop.*, *11*, xxvii.

Thompson, C. (2011). How khan academy is changing the rules of education. *Wired magazine*, *126*, 1–5.

Toufaily, E., Zalan, T., & Dhaou, S. B. (2021). A framework of blockchain technology adoption: An investigation of challenges and expected value. *Information & Management*, *58*(3), 103444.

Trowler, V. (2010). Student engagement literature review. *The higher education academy*, *11*(1), 1–15.

Tsai, Y.-S., & Gasevic, D. (2017). Learning analytics in higher education—challenges and policies: A review of eight learning analytics policies. *Proceedings of the seventh international learning analytics & knowledge conference*, 233–242.

Tsai, Y.-S., Moreno-Marcos, P. M., Tammets, K., Kollom, K., & Gašević, D. (2018). Sheila policy framework: Informing institutional strategies and policy processes of learning analytics. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, 320–329.

Van Inwegen, E., Adjei, S., Wang, Y., & Heffernan, N. (2015). An analysis of the impact of action order on future performance: The fine-grain action model. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, 320–324.

van Klaveren, C., Kooiman, K., Cornelisz, I., & Meeter, M. (2019). The higher education enrollment decision: Feedback on expected study success and updating behavior. *Journal of Research on Educational Effectiveness*, *12*(1), 67–89.

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, *53*(4), 5–23. https://doi.org/10.1007/BF02504682

Whitelock-Wainwright, A., Gašević, D., & Tejeiro, R. (2017). What do students want?: Towards an instrument for students' evaluation of quality of learning analytics services. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 368–372.

Whyte, A., Nayak, P., & Johnston, J. (2016). Lak16 workshop: Extending ims caliper analytics™ with learning activity profiles. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 490–491.

Wood, G. et al. (2014). Ethereum: A secure decentralised generalised transaction ledger. *Ethereum project yellow paper*, *151*, 1–32.

Yang, C. C., Chen, I. Y., & Ogata, H. (2021). Toward precision education. *Educational Technology & Society*, *24*(1), 152–163.

Zapata-Rivera, J. D., & Katz, I. R. (2014). Keeping your audience in mind: Applying audience analysis to the design of interactive score reports. *Assessment in Education: Principles, Policy & Practice*, *21*(4), 442–463.

Zheng, Z., Xie, S., Dai, H., Chen, X., & Wang, H. (2017). An overview of blockchain technology: Architecture, consensus, and future trends. *2017 IEEE International Congress on Big Data (BigData Congress)*, 557–564.

Zhu, Y., Qin, Y., Zhou, Z., Song, X., Liu, G., & Chu, W. C.-C. (2018). Digital asset management with distributed permission over blockchain and attribute-based access control. *2018 IEEE International Conference on Services Computing (SCC)*, 193–200.