A Blueprint for a Blockchain-Based Architecture to Power a Distributed Network of Tamper-Evident Learning Trace Repositories

Juan Carlos Farah[1], Andrii Vozniuk[1], María Jesús Rodríguez-Triana[1,2], Denis Gillet[1]
[1]École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[2]Tallinn University, Tallinn, Estonia
{juancarlos.farah,andrii.vozniuk,maria.rodrigueztriana,denis.gillet}@epfl.ch

***Abstract*— The need to ensure privacy and data protection in educational contexts is driving a shift towards new ways of securing and managing learning records. Although there are platforms available to store educational activity traces outside of a central repository, no solution currently guarantees that these traces are authentic when they are retrieved for review. This paper presents a blueprint for an architecture that employs blockchain technology to sign and validate learning traces, allowing them to be stored in a distributed network of repositories without diminishing their authenticity. Our proposal puts participants in online learning activities at the center of the design process, granting them the option to store learning traces in a location of their choice. Using smart contracts, stakeholders can retrieve the data, securely share it with third parties and ensure it has not been tampered with, providing a more transparent and reliable source for learning analytics. Nonetheless, a preliminary evaluation found that only 56% of teachers surveyed considered tamper-evident storage a useful feature of a learning trace repository. These results motivate further examination with other end users, such as learning analytics researchers, who may have stricter expectations of authenticity for data used in their practice.**

***Keywords*— blockchain, privacy, education, learning analytics**

I. Introduction

The rise of digital education and learning analytics (LA) has led to a significant increase in the amount of information collected during educational activities [1]. This trend motivates the development of new frameworks to safeguard educational data and the privacy of those concerned. Blockchain technology has been specifically highlighted as having a "potential to provoke major shifts in educational practice", particularly in the way learning records are managed [2].

The blockchain came to prominence with the rise of the Bitcoin cryptocurrency [3], functioning as the distributed ledger on which Bitcoin transactions are recorded. It employs cryptographic functions to create an append-only, tamper-evident log, where items get inscribed only if approved by the majority of participants in the network. Over the past few years, it has been proposed as the core building block of decentralized applications in various sectors where user privacy and data authenticity is paramount, including finance [4], telecommunications [5], and healthcare [6], [7]. In educational settings, research has focused on the blockchain's potential to securely register badges and certificates [8], recording and trading educational data and reputation [9], and powering a platform for examinations [2], [10].

In this paper, we propose a blueprint for a system that uses blockchain technology to validate the authenticity of learning traces stored across multiple locations. A system based on this blueprint would give participants in a learning activity (e.g. students, teachers, institutions) control over where their data is stored without undermining its validity, while also supporting a more fine-grained and direct control over when, how and with whom data is shared. By empowering users with more control over their data, systems based on this blueprint could address issues put forth by data privacy regulations, such as the European General Data Protection Regulation (EU GDPR 2016/679) [11], which introduces a series of requirements regarding the collection and processing of personal data to be met starting May 2018.

To assess the impact that this proposal could have on end users, we conducted a survey of 25 teachers, capturing the extent to which they considered the features derived from our blueprint to be useful in their practice.

II. Motivation and Related Work

As underlined by Pardo and Siemens [1], there are ethical and privacy issues that need to be considered when collecting LA data, such as the responsibility to promote trust among participants and accountability from those with access to that data. Additionally, learning traces are predominantly stored by the systems in which they occur. These systems are often not interoperable, making data management and analyses involving multiple data sources a more complex process [12].

The problem of private data stored distributively has been already faced in the medical sector, where patient data is often dispersed among various healthcare providers, and privacy is of utmost concern [7]. An approach to tackle this challenge is *MedRec*, developed by Azaria et al. at MIT's Media Lab [6]. *MedRec* is a platform that exploits blockchain technology to create a network that manages access permissions for electronic medical records. It allows healthcare providers to store the data in their own databases, but grants a way for patients to retrieve all of their records across providers and share them with doctors or third parties.

A number of applications of the blockchain have also been suggested for educational contexts. Among other proposals, blockchain technology has been advocated as a way of validating academic certificates [8] and to create a tradeable currency out of educational achievements, as is the case of *Kudos* [9]. Others present frameworks, protocols, and tools

that coalesce decentralized resources and learning environments (LEs), without applying the blockchain. These include project ROLE's Interoperability Framework [13], the Connected Analytics Toolkit [12], architectures for integrating *e-portfolios* in distributed LEs [14], and learning trace trackers that focus on open learning platforms [15]. Nevertheless, no system comprehensively provides a way to (1) allow data to be stored outside of a central repository, (2) guarantee that it has not been tampered with when it is retrieved for analysis, and (3) integrate with multiple LEs.

These three limitations illustrate the challenges highlighted by Tapscott and Tapscott [10], who argue that the blockchain can be applied for innovation within higher education, specifically in the management of identity and student records. Within this domain, they put forth three challenges to be addressed: (1) maintaining the privacy and security of data stored, (2) assuring the validity of this information, and (3) rewarding learning done outside of the formal classroom. We map these challenges to three design dimensions for our blueprint: (1) *Data Ownership and Access*, (2) *Data Authenticity* and (3) *Data Integration and Aggregation*, respectively. These dimensions define the main pillars of our blueprint.

III. Design Considerations

Our objective was to devise a distributed data store for learning traces, exploiting blockchain technology to ensure the privacy, authenticity, and accessibility of these records. The guiding concern was that by giving users full control over their data, they might be able to manually modify their activity traces and thus rewrite their learning histories. In order to avoid this, our approach was to apply the blockchain as an immutable record to prevent educational data from losing authenticity even when stored in a location controlled by an untrusted third party. Centered on this concern, we mapped the functionalities proposed to the challenges underlined by Tapscott and Tapscott [10], and grouped them under the following three design requirements and approaches.

1. Data Ownership and Access.

Requirement: Learners, parents, teachers, institutions and platform providers should all have a say in deciding where the data is stored. Access permissions need to be signed by the parties involved and could be temporary [11], meaning stakeholders should be able to revoke access to the data.

Proposed Approach: Permissions are registered, updated and verified on the blockchain. Repositories are pluggable data stores that can be disconnected, relocated and reconnected. Repository owners can override permissions locally.

2. Data Authenticity.

Requirement: Participants should be able to exploit their data, validate its authenticity, and share it confidently, even if the institution that hosted the activity or the platform where the data was generated is no longer available.

Proposed Approach: Raw data is not stored on the blockchain, only instructions to retrieve and validate records. Anyone holding raw data can validate it independently.

3. Data Integration and Aggregation.

Requirement: Data from multiple sources — hosted both within and outside of traditional LEs — should be supported.

Proposed Approach: Records are stored in a way that allows aggregation across repositories. Authorized third parties can integrate records reliably. Coordination is abstracted through an application programming interface (API).

IV. Architecture

In this section, we outline the principal processes and components of our blueprint, as well as its overall layout.

A. Overview

Capturing. Central to this architecture is the concept of a *learning activity* that a learner performs within a digital LE by interacting with other users, resources, and services. For learners to take part in a learning activity, they need to *enroll*. A single enrollment can be valid for multiple learning activities, comparable to taking a course. The learning activities we consider for our design are scoped in time and generate *learning traces*, each of which describes an interaction. We propose that at the end of a learning activity, the set of all learning traces for a given learner constitute a *learning block* (LB). A learning block is self-describing, as it contains metadata about the learning activity, such as the LE, the learner, resources or services exploited, and other users involved.

Recording. Figure 1 presents the process through which a learning block gets recorded. Once the learning block has been generated by the LE (Step 1), it is approved first by the learner, who signs it and optionally sends it to other participants (e.g. teachers, tutors, peers) as an encrypted message (Step 2). This process can be automated and is analogous to *submitting* an assignment, with only the intended recipients being able to view the content of the submission. Other participants can then decrypt the submission, verify its content, approve it with a signature, and resubmit until all required parties have signed (Step 3). The hash of a signed *learning block* is then recorded on the blockchain (Step 4) so that it can later serve to verify that the data inside the block has not been tampered with. The block itself is sent to one or more *learning block repositories* (LBRs) (Step 5), which are databases adapted for the storage of learning blocks. Depending on settings determined upon enrollment, LBRs can be hosted by the LE provider, an institution, one or more learners, and/or other third parties. The *owner* of each repository manages the access permissions to the data.

Validating. To retrieve a learning block, the requesting party requires access granted by the owner of the repository. Access permissions are set upon enrollment, but can be amended and even overridden at the repository level by the owner. Hence, the repository owner can always disconnect the repository from the network or change the access permissions locally without requiring the consent of anyone else in the network. If a request is legitimate and the repository is online, a block is verified by comparing its hash to the one recorded on the blockchain and returned if valid.

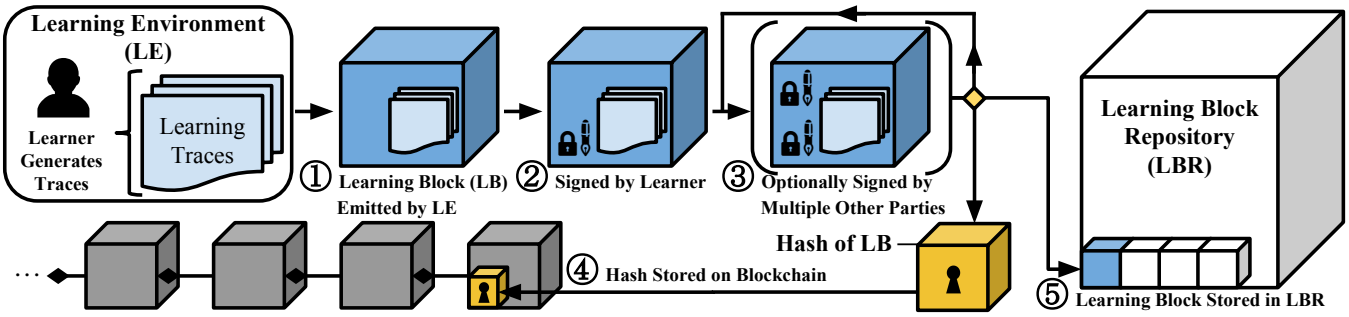

Fig. 1. A learner's interaction with a LE generates *learning traces*, which are then put together and emitted as a *learning block* at the end of the *learning activity*. This block is then signed and sent to an external LBR, with its hash being recorded on the blockchain for future validation.

B. Components

Our proposed blueprint consists of a *blockchain layer*, an *application layer*, and a *communication layer*, as depicted in Figure 2. The blockchain layer comprises the smart contracts that handle the main operations presented in this proposal. The application layer includes components that are completely separate from the blockchain functionality, such as LEs, LBRs, and services performing LA. The communication layer includes all of the modules that handle interactions between the blockchain and the rest of the system, including the end user.

Blockchain Layer. The core of the system will be powered by *smart contracts*, which are an automated way of executing agreements that do not need a trusted third party [16]. We will employ these as a way of managing the locations and access permissions of learning traces, building on the approach proposed by *MedRec* [6]. A principal *enrollment contract* will manage the registration of participants' public keys, network addresses of learning block repositories and other information necessary to sign, validate and fetch learning blocks for a given learning activity or set thereof. It will also *emit* events that other components will listen to in order to relay relevant notifications to end users (e.g. validation, change of address, permission changes). Secondary contracts such as an *access contract*, an *aggregation contract* and a *transcript contract* serve to respectively grant and revoke access permissions, aggregate data across several participants, and provide a single entry point for all the enrollments of an individual participant, among other functionalities.

Application Layer. Components that handle data generation and storage outside of the blockchain are considered to be application modules. These include the *LE*, the *LBR*, and *learning analytics services*. The LE is where enrollment occurs and where the learning activity takes place. It is in charge of hosting the generation of learning traces and grouping them at the end of the learning activity to constitute a *learning block*. The LBR stores the signed learning blocks and the learning analytics service processes data stored across the repositories. It is important to note that only the LBR contains data and that any LA service will fetch the data required for analysis through the communication layer. During this process, data is validated against the blockchain layer. Nevertheless, our blueprint allows for a user with access to the LBR to fetch data directly, without passing through the communication layer or validating the authenticity of the learning blocks. This enables clients that trust the LBR—such as the user or institution that controls it—to read directly from the repository without any additional overhead.

Communication Layer. The communication layer consists of agents that handle the transfer of data and requests between the application modules, the blockchain, and the user. The key idea is to expose an API that is abstracted from a specific blockchain technology, allowing the system to be modular. Agents also *listen* to events emitted by the smart contracts and relay any relevant information to end users or other components. A *Writing Agent* handles the process depicted in Figure 1, namely signing a learning block, registering its hash on the blockchain, sending the learning block to the appropriate repository, and either notifying participants that this transaction was successful or reporting any failures. A *Reading Agent* handles ensuring correct access permissions, fetching and validating a learning block, and sending out success or failure notifications. A third *Registration Agent* processes the initial registration of information and any updates to access permissions, public keys, and addresses.

C. Trust

An important aspect of this blueprint is that the whole communication layer has to be trusted. This requirement emerges due to the fact that it handles all interactions with the blockchain, including the validation of learning blocks and access permissions. A malicious communication layer could fake validation and register spurious records on the blockchain, or allow unauthorized access. Similarly, the blockchain layer has to be trusted, either by using a public blockchain that independently guarantees to be trustworthy, or by relying on a permissioned blockchain controlled by a trusted third-party. Since the contracts registered on blockchains such as Ethereum[1] are immutable, deterministic, open source programs that can be formally verified [17], trust in those contracts can be independently audited. On the other hand, the LBR and the LA services are not trusted, and therefore all of their requests and responses are mediated by the trusted communication layer.

[1]Ethereum (`https://ethereum.org`) is a blockchain application platform that supports Turing-complete smart contracts.

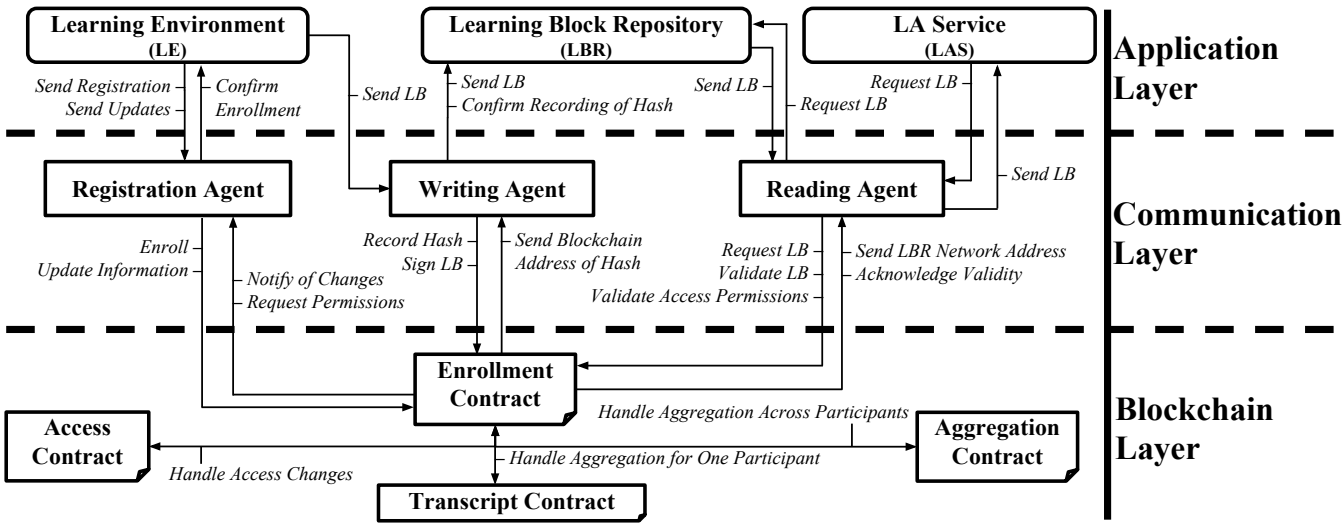

Fig. 2. Our blueprint consists of three layers. Abstracting interaction with the blockchain using a communication layer increases modularity and allows integration with different chains. It also requires that the communication layer be trusted.

V. Preliminary Evaluation

As highlighted in [18], teachers are often the ones deciding what technology they adopt to support their practice. Thus, in order to carry out a preliminary evaluation of the impact an architecture based on our blueprint would have on teachers, we conducted an online survey[2] of 25 teachers with expertise in digital education solutions and previous experience with built-in LA. The aim of this survey was to gauge the perceived usefulness of features that emerge from the **recording** and **validating** processes as described in Section IV-A.

Table I shows the results for four features divided into two groups. Group 1 consists of features 1 and 2, which are enabled by our blueprint, but not directly facilitated by blockchain technology. Group 2 consists of features 3 and 4, which are directly supported in our blueprint through the use of blockchain technology. A discussion of these results follows in the next section.

TABLE I

Teachers who agree that a given feature should be supported by a teacher-owned repository of learning traces. ($n = 25$)

A repository of learning traces should:	Agree
1. Be always accessible to the repository owner.	76%
2. Allow the repository owner to share data.	56%
3. Ensure data has not been tampered with.	56%
4. Verify data was generated by a given student.	36%

VI. Discussion and Conclusions

The blueprint presented in this paper addresses the challenges highlighted in Section II, namely the need to maintain the privacy and security of learning records, assuring their validity, and allowing integration with multiple LEs. Additionally, it provides a solution for systems handling student data to ensure that they are compliant with privacy regulations, such as [11]. Our contribution is the scaffolding for a novel application of the blockchain in education that tackles the privacy requirements of participants, as well as the need for transparency and accountability in the field of LA. We focus on specifying the features required to implement the design considerations outlined in Section III and present a blueprint to create a prototype based on those requirements. Our aim is to provide the foundations for an architecture that allows users to store learning traces in locations of their choice, without sacrificing the ability to guarantee the authenticity of this data.

An architecture based on this blueprint would pave the road for better privacy protection for those participating in online learning activities. Instead of relying on various third parties to continuously store and certify learning achievements, it would allow learners to keep data in one or more repositories that they control, with trust centered around a communication layer that can validate data against a highly-available blockchain. The capability of gathering and integrating records from multiple sources could support personal development and reflective learning across platforms, as proposed in [14] and [19]. Indeed, multiple *e-portfolio* applications could be powered by these trusted repositories. These *e-portfolios* would allow learners to confidently prove attainment of competencies and both reflect on and showcase their achievements, as motivated in [20]. Additionally, trusted repositories of learning traces would grant researchers an opportunity to perform analysis on distributed data with a guarantee that it has not been tampered with, offering more transparency and accountability when conducting LA. Finally, by placing users in control of where the data is located and who can access it, LEs can remove the need to store activity traces themselves, thus addressing certain requirements put forth by [11], such as the storage duration and accessibility constraints described in Article 25.

On the other hand, there are also a number of limitations and potential barriers to adoption that we need to consider when implementing our blueprint. Firstly, placing user data management in the hands of users themselves might result in data being unavailable for LA. Applications dependent on user data will be thus required to handle situations where

[2]Online Survey: `https://bit.ly/2iDLbCZ`

data is missing, providing graceful degradation of services, as well as feedback about the completeness of datasets and the validity of resulting LA. Secondly, given that the blockchain is an immutable ledger, data written to it cannot be modified or removed. Even if this data consists only of hashes of activity traces or public keys, it is important to address any possible ethical and legal requirements, as well as ensuring that users are aware of what data can be erased and what data is permanently on record. Thirdly, as publishing data to a blockchain requires computing power (and possibly fees), we need to ensure that the granularity and frequency of writes to the blockchain are technically and financially viable. Finally, given that blockchain infrastructures are vulnerable to a number of malicious attacks [21], any architecture based on blockchain technology needs to implement safeguards to address these security implications.

Our preliminary evaluation sheds light on the perceived usefulness of the features a blockchain-based architecture could enable, as well as possible roadblocks it could encounter. While 76% of teachers surveyed found that being always accessible was a key feature of a learning trace repository, there were less favorable opinions regarding two features that emerge from our blueprint through the use of a blockchain, namely (a) ensuring data has not been tampered with and (b) verifying data was generated by a given student. These were respectively marked as important features only by 56% and 36% of respondents. Although these results are by no means conclusive, they could be indicative of teachers' indifference to ensuring learning traces are authentic once a learning activity is over. Nevertheless, our findings motivate further evaluation of use cases for a system based on this blueprint, possibly in providing a reliable data source for LA researchers, who may perceive more added value in ensuring the authenticity of data used in their studies.

VII. Future Work

The proposed blueprint is a first step in the design of our architecture and motivates our future work. A proof-of-concept will be developed on the Ethereum blockchain and validated on Graasp[3]. Once we have a proof-of-concept within a single LE, we aim to perform a case study in an educational context with multiple systems to integrate. With this working prototype, we will test the potential to offer cross-platform interoperability, which is one of the problems identified in Section II. Moreover, in order to further assess the potential use cases for a system based on this blueprint, we aim to obtain additional feedback not only from teachers, but also from students and LA researchers. Following this approach, we expect to address the aforementioned potential barriers to adoption, identify our target user base, refine our design and consolidate our architecture in future work.

Acknowledgment

This research has been partially funded by the European Union (grant agreement nos. 731685 and 669074).

[3]Graasp (`https://graasp.eu`) is an online platform used to create learning spaces that generate traces of learner activity for LA usage.

References

[1] A. Pardo and G. Siemens, "Ethical and privacy principles for learning analytics," *British Journal of Educational Technology*, vol. 45, no. 3, pp. 438–450, 2014.

[2] R. d. R. Mike Sharples, M. G. Rebecca Ferguson, Christothea Herodotou, A. K. Elizabeth Koh, P. Chee-Kit Looi, M. McAndrew, Bart Rienties, and L. H. W. Weller, *Innovating Pedagogy 2016*. Institute of Educational Technology, The Open University, 2016.

[3] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 2008.

[4] H. Lycklama à Nijeholt, J. Oudejans, and Z. Erkin, "DecReg: A Framework for Preventing Double-Financing using Blockchain Technology," in *Proceedings of the ACM Workshop on Blockchain, Cryptocurrencies and Contracts*, BCC'17, pp. 29–34, 2017.

[5] A. Hari and T. V. Lakshman, "The Internet Blockchain : A Distributed , Tamper-Resistant Transaction Framework for the Internet," in *Proceedings of the 15th ACM Workshop on Hot Topics in Networks*, pp. 204–210, 2016.

[6] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "MedRec: Using blockchain for medical data access and permission management," in *Proceedings of the 2nd International Conference on Open and Big Data*, OBD'16, pp. 25–30, 2016.

[7] T.-T. Kuo, C.-N. Hsu, and L. Ohno-Machado, "ModelChain: Decentralized Privacy-Preserving Healthcare Predictive Modeling Framework on Private Blockchain Networks," *NIST Workshop on Blockchain & Healthcare*, 2016.

[8] P. Schmidt, "Certificates, Reputation, and the Blockchain," *Medium*, 2015.

[9] M. Sharples and J. Domingue, "The Blockchain and Kudos: A Distributed System for Educational Record, Reputation and Reward," in *Adaptive and Adaptable Learning*, vol. 9891, pp. 490–496, Springer, 2016.

[10] D. Tapscott and A. Tapscott, "The Blockchain Revolution and Higher Education," *EDUCAUSE Review*, pp. 10–24, 2017.

[11] European Union, "Regulation 2016/679 of the European parliament and the Council of the European Union," *Official Journal of the European Communities*, vol. 2014, no. April, pp. 1–88, 2016.

[12] K. Kitto, S. Cross, Z. Waters, and M. Lupton, "Learning Analytics Beyond the LMS: The Connected Learning Analytics Toolkit," in *Proceedings of the 5th International Conference on Learning Analytics And Knowledge*, LAK'15, pp. 11–15, 2015.

[13] S. Govaerts, K. Verbert, D. Dahrendorf, C. Ullrich, M. Schmidt, M. Werkle, A. Chatterjee, A. Nussbaumer, D. Renzel, M. Scheffel, M. Friedrich, J. L. Santos, E. Duval, and E. Law, "Towards responsive open learning environments: The role interoperability framework," in *Towards Ubiquitous Learning*, pp. 125–138, Springer Berlin Heidelberg, 2011.

[14] A. Lozano-Álvarez, J. I. Asensio-Pérez, G. Vega-Gorgojo, and A. Martínez-Monés, "Helping teachers align learning objectives and evidence: Integration of eportfolios in distributed learning environments," *Journal of Universal Computer Science*, vol. 21, no. 8, pp. 1022–1041, 2015.

[15] J. L. Santos, K. Verbert, J. Klerkx, S. Charleer, E. Duval, and S. Ternier, "Tracking data in open learning environments," *Journal of Universal Computer Science*, vol. 21, no. 7, pp. 976–996, 2015.

[16] M. Swan, *Blockchain: Blueprint for a new economy*. O'Reilly, 2015.

[17] K. Bhargavan, A. Delignat-Lavaud, C. Fournet, A. Gollamudi, G. Gonthier, N. Kobeissi, N. Kulatova, A. Rastogi, T. Sibut-Pinote, N. Swamy, and S. Zanella-Béguelin, "Formal verification of smart contracts: Short paper," in *Proceedings of the 2016 ACM Workshop on Programming Languages and Analysis for Security*, PLAS'16, pp. 91–96, ACM, 2016.

[18] M. J. Rodríguez-Triana, A. Martínez-Monés, and S. Villagrá-Sobrino, "Learning analytics in small-scale teacher-led innovations: ethical and data privacy issues," *Journal of Learning Analytics*, vol. 3, no. 1, pp. 43–65, 2016.

[19] L. Stefani, R. Mason, and C. Pegler, *The educational potential of e-portfolios: Supporting personal development and reflective learning*. Routledge, 2007.

[20] H. Barrett, "Balancing the two faces of eportfolios," *Educação, Formação & Tecnologias*, vol. 3, no. 1, pp. 6–14, 2010.

[21] D. K. Tosh, S. Shetty, X. Liang, C. A. Kamhoua, K. A. Kwiat, and L. Njilla, "Security implications of blockchain cloud with analysis of block withholding attack," in *Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, CCGrid'17, pp. 458–467, IEEE Press, 2017.