Kratos: A secure, authenticated and publicly verifiable system for educational data using the blockchain

Dr. Velislava Hillman
Kratos-Education
Cambridge, USA
vhillman@kratos-education.org

Varunram Ganesh
Digital Currency Initiative
MIT Media Lab
Cambridge, USA
ganeshv@mit.edu

***Abstract*—Growing interest in educational data mining (EDM) and learning analytics (LA) to leverage big data and to benefit education and the science of learning has made data ownership an important focus point for institutions and students. While EDM and LA can provide important information that help enhance the quality of teaching and learning, it has become critical to ensure data privacy and student agency over data. In this paper, we introduce Kratos: an immutable and publicly verifiable data management system that enables EDM and LA, while maintaining data privacy and empowering students with a user interface for data governance and participation in school processes. The system aims to achieve data interoperability, which facilitates EDM and LA as incentives to educational stakeholders (policy makers, educators, developers of education technologies, etc.), while prioritizing student agency over their data. Our system gives students and schools an immutable log along with comprehensive access to data that is otherwise scattered across systems and vendors. The underlying set of rules of the system are defined in a set of smart contracts, codified from existing non-virtual agreements [1] between schools and education technology (edutech) vendors. We propose the smart contracts to be deployed on a public blockchain (like Ethereum or Bitcoin), for notarizing and time-stamping various interactions which users of Kratos may have with data. Third parties requesting access to school data have a unique virtual token assigned to them on the blockchain which helps keep track of data modifications, access and use.**

***Index Terms*—educational data mining, learning analytics, data privacy, blockchain, distributed data management systems**

I. Introduction

To improve work and provide high-quality education, school practitioners need interoperable data about students and school processes that exist in various systems and across various education technology providers [10]. Interoperability among different systems is defined as the ability to communicate with and use the functionality of different peer systems and that there is an established seamless communication flow and exchange of information among cooperative systems [30]. In increasingly digitized school environments with various student and operational information systems, autonomous content and data management systems of third party providers of educational technologies such interoperability of data sharing and functionality becomes challenging. In turn, interoperability challenges limit the ability of education stakeholders to successfully evaluate school effectiveness [11], detect and assess learning problems, develop timely interventions, and develop better tools for learning, instruction and assessment [24].

Although open educational data standards exist, generally vendors providing education technologies have no incentive to adopt any single common one, while most also retain primary ownership of the data generated from the use of their products. As a consequence, it becomes hard to monitor and control what vendors collect and how they use student data, while many still have unclear data privacy terms and policies [12]. More recently, a conceptual model for Technology Learning Data Standard or Ed 3.0 has been proposed at the Institute for Electrical and Electronic Engineers but this is yet to gain traction while the official (withdrawn) IEEE 1484.1-2003 standard has not been updated [25].

Schools today use education technology applications in multiple levels - to organize and store longitudinal data, to manage, distribute, and store course content, for student instruction and assessment, and more. Some of the data generated from the use of these applications is stored at district level while other data is managed by vendors in their own servers. This data often becomes inaccessible by teachers, students and school administrators. As a result, data becomes fragmented due to incompatible back-end systems and general inaccessibility. In cases where vendors do share data this often puts the extra burden on teachers to manually insert the new data onto their school system in order to analyse student progress. Some cases exist where offline agreements enable data sharing between school districts and edutech vendors [1]. However, such practices are not done in a seamless and automated manner as data is delivered in the form of 'canned reports', which offer close to zero insights, preventing educators and administrators to develop accurate learning models. This lack of data interoperability and fragmentation stifles any efforts to

978-1-7281-0858-2/19$31.00 ©2019 IEEE

have a comprehensive understanding and utilization of data.

In the United States, the Family Educational Rights and Privacy Act of 1974 (FERPA) is a federal law that was introduced to allow parents the right to have access to their childrens education records, the right to seek to have the records amended and the right to have some control over the disclosure of personally identifiable information from the education records [26]. FERPA also dictates that when a student turns 18 years old or enters a post-secondary institution at any age, the rights under FERPA transfer from the parents to the student. While such provisions are provided under FERPA, the process to transfer said data requires filling in forms and around 45 days [27] for records to be released for review. Due to the lack of a unified data standard, schools are wary on how they would compile these records, while lack of data literacy can pose further limitations as to what data records young individuals and families can request to have [13].

In this work we propose the development of a system for student data management, privacy, accountability and auditability. We take inspiration from existing data standards and construct a common data schema for otherwise disparate data across different edutech vendors and school systems. We seek to enable data transparency and accountability about its access and use through network permissioning and proofs of ownership on a distributed ledger, which we seek to integrate with existing data standards by designing data analytic models. We look at a prototype user interface that can give students, parents and schools access to otherwise scattered and disparate data.

We initially propose this system to be applied in the context of K12 - kindergarten through 12^{th} grade - schools. Our broader goal is to set up a unique decentralized data management system that provides students with full control and visibility of their data collected throughout their educational life. We acknowledge the legal, structural and organizational complexities involving school data of minors and therefore do not delve into these particular subjects. However, we do focus on conceptualizing how a safe space for student data can be created, where the value to education stakeholders is expressed in their ability to comprehensively access and use data for EDM and LA and to students - by engendering a new culture of data literacy and data-driven decision-making.

II. Existing problems and the need for a solution

To contextualize the complexity of problems related to school data and the challenges schools face as a result of the growing digitization of academic processes, we partnered with Cambridge Public School District (CPSD) that administers public elementary and high schools in Cambridge, Massachusetts. With the help of CPSD's Information and Communication Technology Services (ICT) division, we explored the following issues pertaining to data interoperability and took their suggestions on potential mitigation schemes:

- **Data access** - CPSD has agreements with over 100 vendors [1] providing education technologies. Many of them do not provide direct or comprehensive access to data collected when students use their products and services. When data access is available, it is either provided in the form of reports - a summary of information, which the district database administrator can request and download or is made available only to teachers (and in some cases, students) in the form of digital dashboards. When teachers obtain data about their students, additional work is required from them to upload this newly available data [29] into existing school systems and convert it into meaningful information that can help improve their work [13].
- **Lack of compliance with data standards** - School data frameworks vary across districts and states [9] and edutech vendors use different data formats and schema to organize and store student data since they have no incentive to comply with a given standard. The resulting data, which is scattered across different systems poses challenges to educators and other stakeholders who want to assess the impact of education technologies on instruction, pedagogy and learning in order to identify best practices. The lack of incentive to comply with data standards foreshadows any attempt at achieving data transparency and sound and ethical terms of use.
- **Lack of transparency** - The scale, complexity and number of vendors pose challenges for schools to have a comprehensive list of the data that is generated by students. The lack of a list propels vendors to circumvent regulations such as FERPA and results in misuse of student data. Stakeholders in the education sector have been making tremendous efforts to ensure data transparency in order to better achieve data privacy but the effectiveness of these efforts remains to be realized.
- **Security** - Due to the lack of transparency on how edutech vendors organize and structure their data [12], it becomes difficult to understand the different means they employ to ensure data security and best practices of use. In some cases, security is in the hands of third parties, which makes it practically impossible to oversee what other terms of use might follow.

From the above, we infer that schools grapple with ongoing challenges that preclude them from having seamless and comprehensive access to student data, which ultimately diminish further opportunities from deploying EDM and LA. Additionally, these challenges prevent students from accessing and understanding what data is being collected on them, by whom, for what purpose and in what context.

Interoperability challenges between vendors and schools at local or district level pose barriers to cohesive data management and sharing while lacking basic technological privacy infrastructure and accountability [12]. While digitizing and collecting student data is not a new concept [5], technological advances like cloud computing and the Internet of Things (IoT) amplify concerns about data storage and analysis due to their lack of transparency [4].

The problems of fragmented data access in schools, the lack of student data literacy and auditability and increasing

concerns over how such data is being used [2] present a real problem that warrants for a solution. And while the technical solutions proposed by Kratos are not new, we believe this is the first time that advancements from multiple fields have been combined to form a single system, which can act as a guideline for students, schools and vendors; as a step closer into changing the current ecosystem of fragmented data systems across vendors and servers into a network of seamless data sharing and accountability.

III. Architecture

A. Overview

While designing the architecture of Kratos, we had to look at examples of scalable and secure systems that have been functioning as intended by design. We thus modelled the architecture of Kratos upon the Internet, with each critical component as a part of a bigger stack. At a high level, the Internet as it stands today can be theorized as having two major components. The first consists of web browsers, which users use to interact with and access websites. The second comprises a set of standards governing the propagation of data around the Internet like TCP/IP, DNS and more. While different web browsers like Mozilla Firefox, Google Chrome or Opera exist, there is only one set of standards that define how these browsers can interface with the Internet. Indeed, having multiple standards would cause confusion with regards to which standards are used by which parties.

We draw a parallel with the architecture of the Internet as we envision Kratos similarly to comprise two principle components. The first is the Kratos web interface, which teachers, students and parents can use to interact with. We abstract away the complexity of data architecture and smart contracts for data use auditability and instead lead users to focus on the usability of school data as valuable information about school processes, student learning and well-being. The second part of Kratos consists of a set of standards and rules, which define how data is to be accessed, how data access is to be recorded and so on. Other web interfaces similar to Kratos may come about with their own sets of standards, which will do nothing to resolve the problem of data interoperability. Therefore, we believe that consensus on the set of rules among multiple stakeholders is critical to enforce.

To guarantee that all participants who enter into a network for data sharing and exchange follow stipulatory terms and conditions is to formalize the legal agreements that are otherwise signed between schools and third-party providers and stakeholders. In Kratos, we do that in the form of smart contracts, which are codified representations of the non-virtual agreements [1] that are already in place between schools and edutech vendors (as is in our case study at CPSD). These smart contracts define the rules and conditions for data access and use. Formalizing the above in a smart contract gives both schools and vendors the opportunity to explore parts, which may not be clear to either of them and give suggestions on how some of these might be improved.

Taking the analogy made with the architecture of the Internet, the Kratos platform is like a web browser - an interface where users can interact with their data. The smart contract makes the use of a blockchain for notarizing and time-stamping various interactions, which users of Kratos have with their data. Each vendor has a unique virtual token assigned to them, which they can use to interact with the smart contract.

The data itself is stored encrypted on a distributed file storage system to ensure data availability and consistency. The encryption key would belong to the school and above the age of 18 - to students. When a user requests data access, the smart contract decrypts parts of the data and returns it to the user. In parallel, the smart contract commits both the interaction of the user requesting data access with the smart contract and the time at which the request was made onto the blockchain. This time-stamping mechanism provides the ability to audit and track data ownership and data access requests.

B. Platform

The primary goal of the web platform is to be user-friendly and accessible since Kratos aims to reach a wide range of users - parents, educators and young people. In order to achieve this, we studied various popular learning and social media applications targeting young users to identify clean designs and minimalist user interfaces. We also showed the prototype to a small subset of students and gained their feedback on how various functions and designs can be improved and added in a live chat option based on participants' inputs. The ability to discuss with other students in real time is crucial to arrive at solutions to problems and the feeling of togetherness that students get while interacting with other students is a bonus.

The prototype platform classifies data into two types - Static data and Dynamic data. Static data refers to those data that do not change with progression in time like date of birth or blood group. This data is usually filled in one by the school administrators and then stored for future reference. Dynamic data refers to data that changes frequently like grades or attendance. This data is filled in at frequent intervals by teachers or school administrators and is updated periodically in the school database system. Apart from this, the platform also classifies data on the basis of subject and vendor, which gives easy break-down of the kind of data vendors access and also generate as students interact with their educational products. This particular functionality achieves convenient auditing when a student or parent wants to curb access to specific sections of data, want to view the data points that make up a final grade or identify the source of such data.

The platform hosts a dictionary which contain definitions of otherwise uncommon terms like 'static', 'dynamic' or 'aggregated' data, 'adaptive' learning technologies, data schemas and so on. The dictionary is useful for parents, students and teachers to know more about the classifications of data being collected and where such data might be potentially put to use. Finally, the platform has a comprehensive list of all edutech vendors a specific school partners with and thus aims to create a culture of transparency where stu-

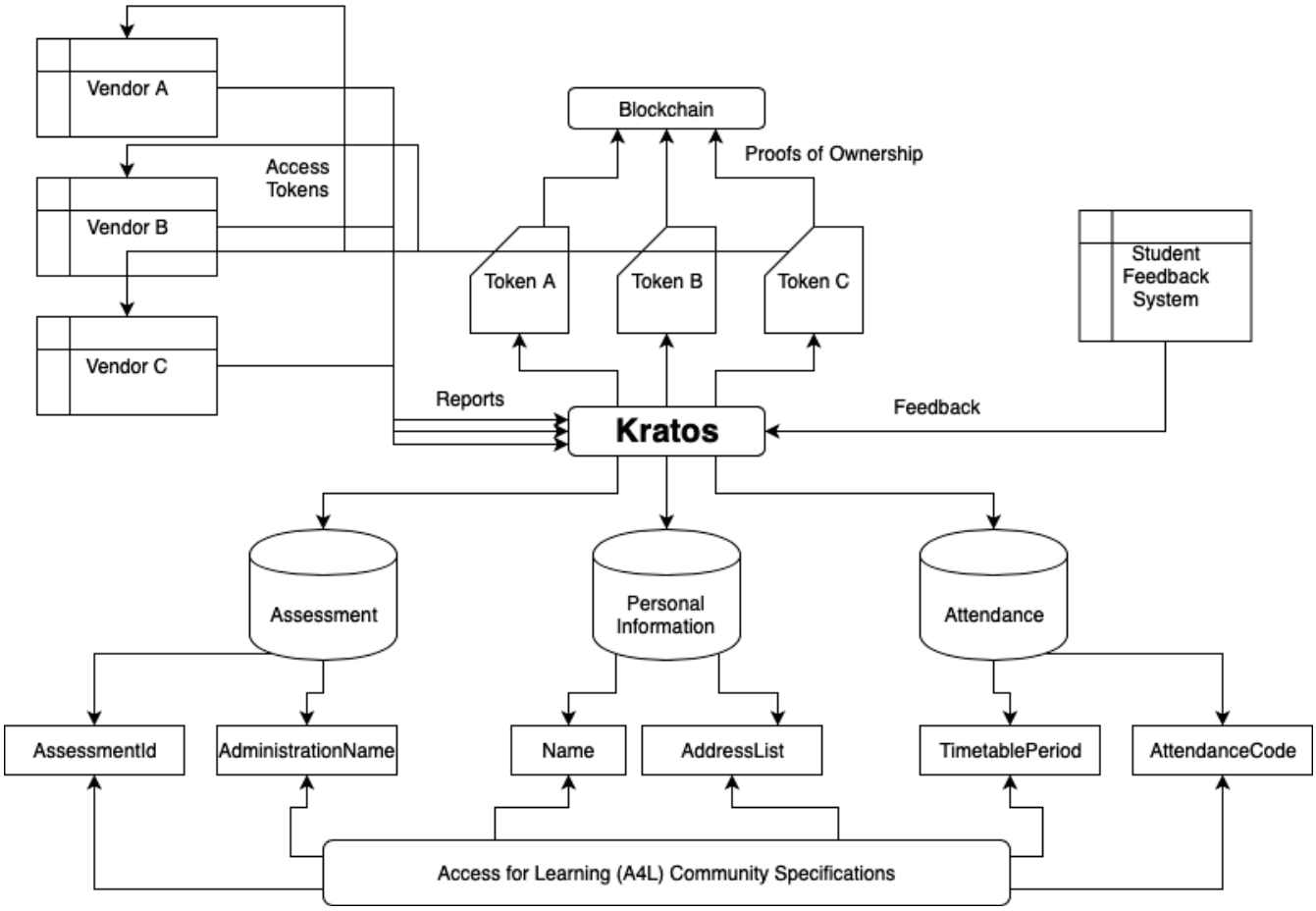

Fig. 1. Kratos Architecture

dents, parents and educators can learn about all participants in a school ecosystem. Additionally, this function is useful for school administrators to swiftly navigate through vendor agreements and monitor and control their data access and use. Additionally, many vendors merge or continue to grow their acquisitions crossing industries and their plethora of data. For example, Alphabet, the company that owns Google and G-Suite Education applications used in schools, has over 200 acquisitions across various business sectors [30]. This raises concerns about the sheer volumes of data that individual for-profit corporations collect from a wide spectrum of services and sectors they own. Kratos's dictionary and interactive functionality creates consistent awareness regarding the role of data in education and allows students to make enquiries directly with their schools and vendors about various aspects of their data and data-driven decision-making surrounding their learning assessments. We prioritize these functions with the aim to increase digital and data literacy opportunities for children and young people assuming an applied learning methodology [36], whereby students learn about the science of data (EDM and LA) - limitations as well as advantages - by observing and interacting with their own school data.

C. Blockchain

The blockchain is one of the most interesting innovations that has happened during the past decade. Blockchains have made it possible to have a secure and immutable time-stamping system, along with the functionality of deploying complex smart contracts for people to publicly verify and audit contract execution. In the recent past, there have been multiple blockchains focused on specific applications like climate change or donations and this variety in choice offers us the ability to be able to choose and integrate multiple blockchains depending on the use case.

In review of the functionalities blockchains enable and Kratos's requirements, we primarily see three requirements: Kratos needs a secure and immutable time-stamping system in order to verify data access at different intervals. We acknowledge the need for public verification of smart contracts to ensure data use auditability and we require that this functionality is available at a low cost since schools and third-party vendors might not be able to afford frequent, multiple and costly interactions with the blockchain. As a footnote, we observe that the blockchain layer we adopt is secure against hackers and other malicious entities to preserve data integrity

and validity of proofs. In other words, we require a blockchain that has withstood the test of time against malicious work.

The Ethereum blockchain offers a stable solution to our needs described above. The Ethereum Virtual Machine (or EVM) allows for execution of Turing complete smart contracts on the blockchain and for granular time-stamping of data with the possibility of time-stamping interactions every 15 seconds. Ethereum also has the biggest and most vibrant developer community in the world with experienced people from multiple domains actively contributing to the ecosystem. But Ethereum has a set of disadvantages, too. Its transaction pricing market (referred to as "gas market") is prone to volatility[37], resulting in sudden transaction fee spikes. The "gas", which is the underlying asset used to power these transactions has a block level limit and any transaction that is unable to pay the required amount of gas is not propagated to other peers. Finally, Ethereum itself is expensive, averaging $200 for one unit of Ether and given transaction costs are denominated in Ether, this increases the overall cost of transacting on the Ethereum blockchain. The Stellar blockchain also offers a solution to Kratos's needs described above. Stellar is a blockchain focused on cross border payments and is designed with the intention of keeping transaction costs low. Lumens themselves are inexpensive, averaging at $0.08 per lumen and the protocol parameters fix transaction costs at 0.00001XLM [20], making day-to-day transactions inexpensive. Stellar also offers better granularity in time-stamping, with the possibility of time-stamping interactions every 5 seconds on average. Stellar is also more energy-friendly than Ethereum, not relying on expensive Proof-of-Work computations but instead relying on its Federated Byzantine Fault Tolerance (FBFT) consensus algorithm [19]. Stellar does not offer the possibility of public verifiability of smart contract execution, and does not offer the ability to execute complex non-financial transactions. This affects auditability, since contract computation is not publicly verifiable and affects application versatility as we are forced to translate all interactions in a strictly financial sense.

Both Ethereum and Stellar have their advantages and disadvantages and since Kratos is purely application-oriented, the design of the system should ideally be modular and blockchain-agnostic. For the purpose of the pilot prototype however, we chose to go with Stellar because we wanted to frequently time-stamp interactions and gain feedback from education stakeholders and given we were operating in a trusted environment with CPSD, we relaxed the assumptions around public verifiability. For the future however, we will be developing smart contracts and deploying them on the Ethereum blockchain to critically evaluate the advantages offered by both in a production environment.

D. Smart Contract

A smart contract is a set of promises, specified in digital form, including protocols within which the parties perform on these promises [22]. In the context of Kratos, a smart contract is a codified representation of the physical agreements that are currently in place between schools and vendors [1].

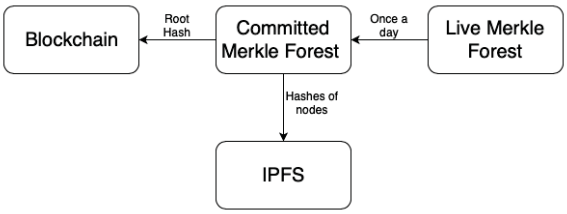

Fig. 2. Data Storage

Kratos defines these smart contracts in order to formalize the notions of data privacy and access, and quantify the risks of not pertaining to them.

The smart contract can either be deployed on a public blockchain system like Ethereum or can be deployed on a centralized hosting provider with proofs of contract execution being committed regularly to the blockchain. In the context of the Stellar blockchain, the latter is the only choice since Stellar does not offer the option of deploying these contracts in a public fashion [23]. Generating proofs of contract execution in such scenarios is tricky and we describe the structure that we have designed in the context of Kratos in the following section.

In the context of Ethereum smart contracts, these can be easily deployed on-chain and batched proofs of data access can be stored as state variables directly on the blockchain. There are multiple optimizations around gas costs, transaction fees and so on, which would again be a topic of focus once we move past the pilot prototype.

E. Data Access and Retrieval

Data Access Verifiability and Auditability is one of the major design goals of Kratos. The platform provides proof that certain data was accessed at a certain instance in time and has an auditable log of interactions with data. This key functionality is achieved by using blockchain based "assets" or "tokens", which vendors use to request and access data. This structure makes use of a forest of Merkle trees whose structure we propose below.

We note here that illegally making copies of data which vendors have access to can not be prevented by Kratos or any other data system. In the context of schools, the Children's Online Privacy Protection Act (COPPA) provides an existing base layer legal system [12] onto which Kratos takes to enforce penalties and prevent out-of-band data storage.

The structure that we use to prove data access and have an auditable log is tricky to design with multiple constraints around latency of the underlying blockchain layer and the frequency of updates made to the data structure. Committing data to the blockchain too frequently results in compounding costs and committing data at very few intervals results in a lack of proof availability for an extended amount of time. We seek to achieve a middle ground between both, and thus have identified that a day's granularity is satisfactory for most educational institutions and vendors.

Fig. 3. Constructing the Merkle Forest

The frequency of updates made to the underlying data structure, however, is independent of the blockchain layer. Updates need to be made immediately and adding latency to this operation results in poor user experience. Our new structure is built on top of existing work of hash-based accumulators [35] and optimized for Kratos's application for storing student data.

1) Accumulators: Accumulators are compact representations of a set, to which elements can be added and proven. Our accumulator, which builds upon the work of [35] uses a forest of Merkle trees. We design our accumulator to allow for efficient add, delete and modify operations and augment the tree nodes to store the number of elements below them to allow for easy membership count queries and potentially other operations.

2) Design of the Accumulator: At the primary level, we use a collection of Merkle trees, which are organized in a specific way. We construct a binary tree for each student in Kratos, independent of their school or class, with the leaves acting as the data points and nodes acting as the data identifiers. Since there are a multitude of data identifiers, we sort them alphabetically and only allow two nodes to be at a single level, thus preserving the structure of a binary tree. As in standard Merkle trees, we hash the students below a specific node all the way up to the root of the tree, and we call the hash of the student tree root as "tree hash". The student trees are then joined together and classified on the basis of a student class (Classes K-12). If there are 15 classes or fewer in a K-12 school, there would be 15 hashes denoting the accumulation of the data present in each of these classes. We define the hash of a particular class as "root hash" and we define the hash of the concatenation of all the root hashes as the "forest hash", which is a unique identifier that can be used to identify the school in question.

3) Adding and removing elements to and from the accumulator: Addition is of two types - adding schema associated with a particular student and adding a new student. Adding new schema to a tree is similar to adding a node to a Merkle Tree. First, we traverse the tree to find the alphabetical neighbours of the node using the hash identifiers and then proceed to add the node and update the nodes that are in path up to the root node of the particular student. Adding a new student involves constructing a new Merkle tree and proceeding to link this tree as a sub-tree to an already existing class tree.

Similar to addition, deletion is of two types, too. Deleting a student is as simple as traversing the class tree to find the student deleting the tree and updating the hashes on the path up to the root node of the class tree (and subsequently, the forest hash would be updated). Deleting a specific node is similar to normal binary tree deletion and would mean deleting all the node's students and updating the hashes on the way up to the root node of the class tree.

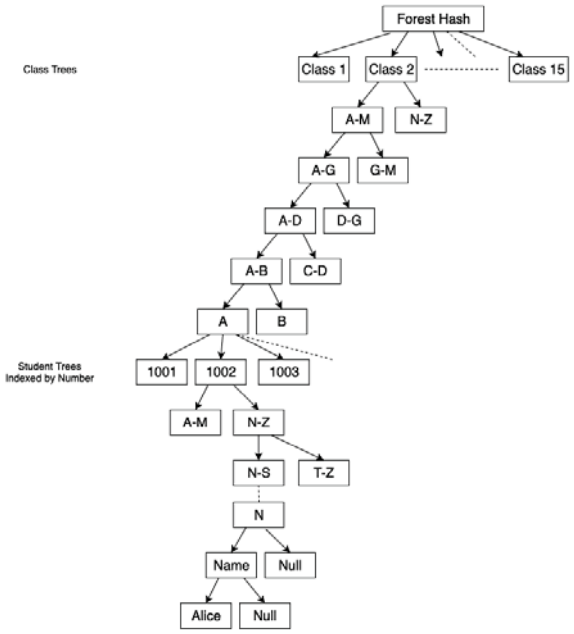

Fig. 4. Merkle Forest Structure

4) Committing data to the blockchain: Since the forest hash is a unique identifier that can be used to reference an entire school, we use it as a proof of data change and commit it to the blockchain. But given data can change multiple times within an hour, committing the hash to the blockchain whenever a piece of data changes is not economically feasible. Instead, we opt for a slightly conservative approach which takes into account the degree of freshness that schools and vendors might expect from student data and commit the forest hash once a day to the blockchain. In parallel, we commit the tree hashes to IPFS and publish its hash publicly on the Kratos platform.

We note that these parameters act as a lever and are not fixed for perpetuity; they can be changed or tweaked by schools in future implementations of Kratos.

5) Accessing data: Data access in Kratos broadly involves two steps. The first is for a requesting entity to present an access token to the Kratos smart contract and verify whether this entity indeed has access to the data it requests. The second is to give the data itself from storage, along with proofs (the hashes on the path up to the forest hash) in order for the verifying entity to know that the data given by the smart contract was not tampered with. Since the hashes are committed only once a day to the blockchain, the data returned by the smart contract would be at most old by a day. This would also mean that the storage layer store two copies of data - the first, a 'read only' structure, which contains the data that was last committed to the blockchain; the second -

a 'write only' structure, which constantly keeps track of the changes made to the data in question. An unrelated benefit of this split is that read and write access are isolated, reducing chances for a user to tamper with the accessible data.

F. Data Storage

Kratos uses a combination of two data storage layers, each serving their own purpose. The first layer that we use for storing the two Merkle forests that we described above is postgreSQL. PostgreSQL is a production-ready SQL environment which can handle high latency volumes. Since we do not anticipate Kratos testing the upper performance limits of postgreSQL, it provides us with an easy, out-of-the-box, production ready option that we can use without having to worry about database optimization.

The second layer we use is the Interplanetary File System (or IPFS) that we use as a place for storing the node hashes that are committed on a daily basis to the blockchain. IPFS is designed with scale in mind and provides easy APIs that can be used to interact with the underlying distributed network.

G. Data Privacy

Data privacy remains an ongoing concern for all stakeholders in the educational ecosystem. While laws such as COPPA provide a legal baseline for protecting student privacy, many edutech vendors remain with unclear terms and conditions of data governance and use [12]. Cases of data leaks, hacks and compromised student data privacy continue [32]. Some [33] argue that the current cybersecurity protocols are inefficient to secure student data, while others insist that school administrators and personnel must undergo data privacy and security training [34], which puts further financial stress on schools.

As data privacy and security remain a recurrent problem in the education sector, Kratos mitigates the risk of data theft and storage through the deployment of IPFS for encrypting data and enabling decryption only via the platform, while the use of standard Key Management Solutions (KMSes) further reduces the possibility of data theft. The combination of postgreSQL and IPFS improves the overall speed, reliability and the ability to retrieve data by combining the advantages offered by both systems.

Enforcing transparent terms and conditions of data governance and exchange through Kratos's design, time-stamping any data sharing and exchange commitments on the blockchain and providing clarity about these processes through a simple web interface create a transparent and accountable ecosystem, which prioritizes student well-being and improved education.

IV. Data literacy

A. Implications of data-driven learning environments

In an increasingly data-driven society, it becomes imperative not only to understand data but to interact and participate in the decision-making processes that it begins to impact. While EDM and LA provide insight about how systems operate, the inferences drawn from data can be without context. Fine-grained data collected over long periods of time destroys privacy and can lead to negative impact on individual well-being and pose limitations over future opportunities [7]. The growing digitization of schools creates learning environments that enable constant incoming data streams, the access and use of which become hard to control and audit.

B. Awareness

Being aware of what data is collected and by whom and learning about the implications, the limitations and the benefits from data mining and learning analytics are necessary steps societies must take in our greater effort to assess and understand the meaning of data-driven futures. Kratos aims to provide a common interface for students, teachers and vendors alike to interact with data and better understand its role and use.

V. Conclusion and future work

In this paper we presented Kratos, a system for data accountability, auditability and transparency. We explored the need for a concrete data management solution and how Kratos would fit in with a technical system to interface between multiple disconnected data systems across multiple educational stakeholders. We demonstrated our concept, which is upon a case study of the Cambridge Public School District's system and the challenges they face. We described the growing issues facing schools with regards to school data privacy, data use and the need for data literacy and education for students - concerns that directed our focus on usability and applied data learning. We acknowledged the importance of EDM and LA to various educational stakeholders and presented a system that can enable data sharing and exchange in an immutable and transparent manner.

We look forward to working with academic institutions and schools to understand different roadblocks in ensuring vendor compliance to laws and regulations. We consider investigating how Kratos can enable EDM, LA, student data learning and agency in more protective jurisdictions (e.g. in the context of Europe and its General Data Protection Regulation). We intend to pilot the Kratos prototype in new partnering schools (both in the US and within the European Union) to gain feedback and make improvements. We plan to carry out case studies with end-users such as students, parents and teachers to assess the usability of the system and to optimize functionality and design. We maintain our goal to make Kratos an open source system, which will impose clear data governance for every participant on the network whereby our commitment remains to giving students agency and better understanding of the role of big data in their lives in school and beyond.

Acknowledgment

We thank Cambridge Public School district, Student Data Privacy Consortium, and Access for Learning (SIF) for their support, constant, and timely feedback.

VI. BIBLIOGRAPHY

1) Cambridge Public School District Agreements Listing (2019). Retrieved from: https://sdpc.a4l.org/district_listing.php?districtID=457
2) Zeide, E. (2017). 19 times data analysis empowered students and schools: Which students succeed and why? Retrieved from: https://ssrn.com/abstract=2754438
3) Cody, A. (2013). Will the data warehouse become every student and teacher's 'permanent record'? *Education Week*, May 20.
4) Sultan, N. (2010). Cloud computing for education: a new dawn? 30 *International Journal of Information Management 109*.
5) Fitzgerald, B. (2014). Data collection isn't new. And it predates common core. *Funny Monkey.*
6) Gibson, D. C., Webb, M. E., and Ifenthaler, D. (2019). Measurement Challenges of Interactive Educational Assessment. In *Learning Technologies for Transforming Large-Scale Teaching, Learning, and Assessment*, 19-33. Springer, Cham.
7) Altman, M., Wood, A. B., O'Brien, D., and Gasser, U. (2018). Practical approaches to big data privacy over time. *International Data Privacy Law, 8(1)*, 29-51.
8) Bowler, L., Acker, A., Jeng, W., and Chi, Y. (2017). It lives all around us: Aspects of data literacy in teens lives. *Proceedings of the Association for Information Science and Technology, 54(1)*, 27-35. DOI:10.1002/pra2.2017.14505401004
9) Common Education Data Standards - CEDS 101. Retrieved from: https://ceds.ed.gov/pdf/ceds-101.pdf
10) Snyder, T.D., de Brey, C., and Dillow, S.A. (2016). Digest of Education Statistics 2015 (NCES 2016-014). *National Center for Education Statistics, Institute of Education Sciences*, U.S. Department of Education: Washington, DC.
11) Militello, M., Bass, L., Jackson, T. K., and Wang, Y. (2013). How data are used and misused in schools: Perceptions from teachers and principals.*Education Sciences, 3*, 98-120.
12) Kelly, G., Graham, J., and Fitzgerald, B. (2018). *State of edtech privacy report.* Common Sense Privacy Evaluation Initiative. San Francisco, CA: Common Sense.
13) Data Quality Campaign. (2018). *What parents and teachers think about education data*. Retrieved from: https://dataqualitycampaign.org/resource/what-parents-and-teachers-think-about-education-data/
14) Student Data Privacy Consortium. (2018). *Student Data Privacy Consortium: Policy and procedures*. SDPC. Retrieved from: https://privacy.a4l.org/wp-content/uploads/2018/06/Student-Data-Privacy-Consortium-P-and-P-2018-Final.pdf
15) Sirota, D. (2013). Big data means kids' "permanent records" might never be erased. *MOTHERBOARD*, October 24. Retrieved from: http://motherboard.vice.com/blog/permanent-records-are-hurting-kids
16) Olson, P. (2019). Pearson hack exposed details on thousands of U.S. students. *The Wall Street Journal*, July 31. Retrieved from: https://www.wsj.com/articles/pearson-hack-exposed-details-on-thousands-of-u-s-students-11564619001
17) Kaplan, M. (2019). Who should bear the cost of data interoperability in K-12 education? *Edsurge,*Retrieved from: https://www.edsurge.com/news/2019-07-02-who-should-bear-the-cost-of-data-interoperability-in-k-12-education
18) Azaria, A., Ekblaw, A., Vieira, T., and Lippman, A. (2016). MedRec: Using blockchain for medical data access and permission management. In *2016 2nd International Conference on Open and Big Data*, 25-30. IEEE.
19) Mazieres, D. (2015). The stellar consensus protocol: A federated model for internet-level consensus. Stellar Development Foundation, 32.
20) Stellar Development Foundation. Assets. Retrieved from https://www.stellar.org/developers/guides/concepts/fees.html
21) Amazon Web Services. Amazon Compute Service Level Agreement. https://aws.amazon.com/compute/sla/
22) Szabo, N. (1996). Smart Contracts: Building Blocks for Digital Markets. Retrieved from http://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart_contracts_2.html
23) Stellar Development Foundation. Assets. Retrieved from https://www.stellar.org/developers/guides/concepts/assets.html
24) Baker, R., and Inventado, P.S. (2014). Chapter X: Educational Data Mining and Learning Analytics.
25) IEEE 1484.1-2003 - IEEE Standard for Learning Technology - Learning Technology Systems Architecture (2003). Retrieved from https://standards.ieee.org/standard/1484_1-2003.html
26) Family Education and Privacy Act ((34 CFR 99.31) https://studentprivacy.ed.gov/faq/what-ferpa
27) FERPA for Students. (2015, June 26). Retrieved from https://www2.ed.gov/policy/gen/guid/fpco/ferpa/students.html
28) Kaplan, M. (2019). Who should bear the cost of data interoperability in K12 education? *EdSurge*, July 2
29) Merkle tree. (2019, November 2). Retrieved from https://en.wikipedia.org/wiki/Merkle_tree
30) Bill & Melinda Gates Foundation (2015).*Teachers know best: Making data work for teachers and students*. Bill & Melinda Gates Foundation.
31) Wikipedia (no date). List of mergers and acquisitions by Alphabet. https://en.wikipedia.org/wiki/List_of_mergers_and_acquisitions_by_Alphabet
32) Cameron, D. (2017). 1.3 Million U.S. students exposed in now-secured data breach. *Daily Dot*,

April 20, https://www.dailydot.com/layer8/1-3-million-american-students-exposed-data-breach-now-secured/
33) Levine, E. (2018). Three ways K-12 schools can fill overlooked cybersecurity holes.*EdTech Focus on K-12*, August 17, https://edtechmagazine.com/k12/article/2018/08/3-ways-k-12-schools-can-fill-overlooked-cybersecurity-holes
34) Herold, B. (2017). With hacking in headlines, K-12 cybersecurity education gets more attention. *Education Week*, March 21, https://www.edweek.org/ew/articles/2017/03/22/with-hacking-in-headlines-k-12-cybersecurity-ed.html
35) Dryja, Thaddeus. "Utreexo: A dynamic hash-based accumulator optimized for the Bitcoin UTXO set."
36) Ormrod, J.E. (2016). *Human Learning* (7th Ed.). Upper Saddle River, NJ: Pearson.
37) Chan, Wren & Olmsted, Aspen. (2017). Ethereum transaction graph analysis. 498-500. 10.23919/IC-ITST.2017.8356459.