

# The Equity and Inclusion in Higher Education: A Proposed Model for Open Data

Carla Hamida<sup>1</sup>, Amanda Landi<sup>2</sup>, Ziyi Liu<sup>3</sup>

<sup>1</sup>*Bard College at Simon's Rock, Great Barrington, USA (Indonesia), chamida16@simons-rock.edu*

<sup>2</sup>*Bard College at Simon's Rock, Great Barrington, USA, alandi@simons-rock.edu*

<sup>3</sup>*Bard College at Simon's Rock, Great Barrington, USA (China), ziyiliu16@simons-rock.edu*

**ABSTRACT:** Recently, governmental institutions and private industries in power have been pushed to be more transparent so that more people can have ownership of their data. Another type of institution with a large amount of power over data are educational institutions. Colleges and Universities around the globe store a significant amount of data on millions of students, such as financial aid, grades, dropout or graduation, successes after graduation. Each institution is rated with respect to these items and more, and potential students are making decisions to go to the school based on these ratings. Therefore, it is imperative for students, who invest their time and their money into the school of their choice, to know the truth. In 2017, the College Transparency Act and the Student Right to Know Before You Go Act were passed, which were created to push transparency for data in higher education. The openness of data in higher education will be beneficial to prospective students. The push for these two bills coincided with the bitcoin bubble. In the past three years, experts in economics, medicine, and supply chain management have been researching methods on how to implement blockchains to create optimal and decentralized data systems. In this paper, we propose a model for open data in higher education inspired by the Bitcoin, which uses blockchain. When used together with InterPlanetary File System, a peer-to-peer distributed file system, we can create a decentralized platform that increases accessibility of data and autonomy of prospective students.

**KEYWORDS:** open data, higher education, blockchain, IPFS, transparency

## Introduction

In today's society, data is currency. Many stepped into the market by collecting data, e.g. Google, Facebook, National Security Agency. Others, still, monetized *controlling* the access and use of the data, e.g. Facebook, government spending budgets. Open data is defined by the Open Data Institute as data that anyone can access, use, or share (Open Data Institute 2017). Across the globe, nonprofit organizations are pushing for empowering citizens with data. For example, the Open Data Charter was founded in 2015, and it is a collaboration of more than 70 governments, experts, and organizations whose sole goal is to make governmental data more available and accessible to citizens of the world. The Open Data Charter proposed six principles, and they are meant "for improved governance and citizen engagement" and "for inclusive development and innovation" (Open Data Charter 2015). While governments and private for-profit companies certainly play a role in the monitoring and controlling of data publication, education institutions make huge profits from their management of student data.

Transparency and accountability is imperative in higher education. Prospective students need accurate information with respect to financial aid, program success statistics, job- obtained-after-graduation data, demographic statistics, and other forms of cost such as living and food. Educational researchers, accreditation teams, and governments investing financial aid need granulated data on student success so that inclusivity of marginalized groups can be improved (Koch 2018). However, transparency does not mean simply listing summarized data online. In fact, every college or university that receives federal aid from the United States is legally required to submit raw data regarding their demographics and financial aid reports annually (Schneider 2017). This information is available to the public, as it is on the The Integrated Postsecondary Education Data System (IPEDS) website. However, navigating through the website itself is a hassle, and large chunks of data must be downloaded in order to attain the raw data for each institution. We need for these institutions to publish simplified aggregate data in order to fully understand how much change has been made.

In Section 1 of this paper, we expand on why transparency and accessibility is required of higher education institutions. In Section 2, we discuss the major issue of privacy of student data with respect to making data more granular. In Section 3, we explain the current blockchain technology so that in section 4 we can propose our solution to the issue of privacy as a stumbling block to complete transparency of higher education data. Finally, in section 5 we conclude our paper and state paths for future work.

### **Section 1. Transparency in Higher Education**

In 2019, there is still a lack of representation of various people in higher education institutions. Organizational change is slow and it only happens effectively when all members involved in and affected by the change see the value in implementation of the change (Berg and Hanson 2018). Although there have been efforts to inspire this change, such as Affirmative Action which first appeared in the Supreme Court in the 1978 case *Regents of California v. Bakke* and scholarships for underrepresented people in higher education (West 1998), one reason for such little change in the last several decades is the still present sexism, racism, classism, and homophobia among the student body as well as the admissions process.

Another huge reason is the lack of access to data that tracks minority students, providing educational researchers the ability to determine weaknesses in a program and allowing curricular developers the chance to improve courses. In 2017, the STEM Research and Education Effectiveness and Transparency Act was passed. The purpose of the bill is to promote inclusion of marginalized groups, specifically women, in participation of research in STEM. Section 2 Article 2 of the bill emphasizes the need to continually collect “information on student outcomes using all available data, including dropout rates, enrollment in graduate programs, internships or apprenticeships, and employment” so the development of marginalized groups in STEM can be tracked (US House 2017). We often read that universities are becoming more inclusive on the news, e.g. (Association of American Colleges & Universities 2015), (Esters), and (Smith 2018). However, many people either do not have access, or have little access to, the actual data informing the demographics at universities. Moreover, the data available may not break demographics down into specific fields and undergraduate v. graduate programs. Universities such as Harvard and Cambridge publish annual reports on their demographics. Even after these universities publish annual reports, it is still inconvenient for readers to open each annual report to compare the progress between these higher education institutions.

Given the continued existence of institutional marginalization, there is a great need to create and implement new policies and solutions. We need to implement a more optimal allocation of resources that can provide real impact to young lives. Understanding the issues within the higher education system, and how these issues affect students, could be done in a systematic matter if all the information was collected on one network in an accessible manner.

### **Section 2. Transparency v. Privacy: Efforts to Protect Student Data**

Despite our need to publish accurate and granular data, we still need to protect the identities of the students represented in the data (ensure anonymity). While there is concern for the misuse of existing data in higher education, it does not mean that is a reason to abandon the idea of sharing. Rather, it means that we need to build systems and establish unambiguous policies in place to protect the data. In 2017, the College Transparency Act was introduced; the bill requires that the National Center for Education Statistics create a data system that analyzes financial costs and student enrollment patterns, customizes information for users accessing the data system, and have the ability to link with other federal data systems (US Senate 2017).

In addition to the College Transparency Act, in November 2017, the Student Right to Know Before You Go Act was introduced by Senators Marco Rubio, Mark Warner, and Ron Wyden (US House 2017). The purpose of this bill is to publish granular and uncomplicated data on higher education institutions in order for prospective students to create informed decisions when applying to colleges and universities, while maintaining privacy standards. The bill requires the data

platform use encryption technology that “includes the use of secure multi-party computation, which generates statistical data based on information provided by colleges and universities as well as loan and income information from government agencies like the IRS” in order to keep published information anonymous (Ortega 2017).

The push for more open data in higher education has not only come from congress. Private organizations such as Data Quality Campaign have been advocating the need to make student data more accessible since 2005. In 2014, the Data Quality Campaign and the Consortium for School Networking established the “Student Data Principles.” There are 10 principles that promote the openness of data in higher education and the use of data to create inclusion in the academic world.

The technological platforms being used for higher education data need to be advanced enough to meet the needs in the future. As seen with IPEDs, handling student data can be complex since different governmental organizations and schools have unique ways of collecting their data. Given that the data could also be misused, those holding the complete and raw data are responsible to ensure that identities remain anonymous. We need to establish a coherent system in which information can spread across the network and each party has the ability to access the appropriate information while being able to update the network systematically with complete information. Handling a large amount of information often leads to complications with storage. In the past, establishing such a network was a more difficult task. In the present, however, decentralized and transparent data systems have been created and implemented. One such data system we next discuss is blockchain technology.

### Section 3. The Model

#### Section 3.1. Blockchain

Blockchain was originally designed to store Bitcoin transactions (Zheng 2017). At the basic level, it is a list of blocks that contains certain information. Figure 1 illustrates a simple blockchain model.

1. Index: the index of the block
2. Timestamp: the time when the particular block was created
3. BPM: pulse rate, an example of the kind of data that can be stored in a block
4. Hash<sup>1</sup>: a unique hash of the block, which is calculated based on all the information stored in the block
5. PrevHash: the hash of the previous block to link the blocks together

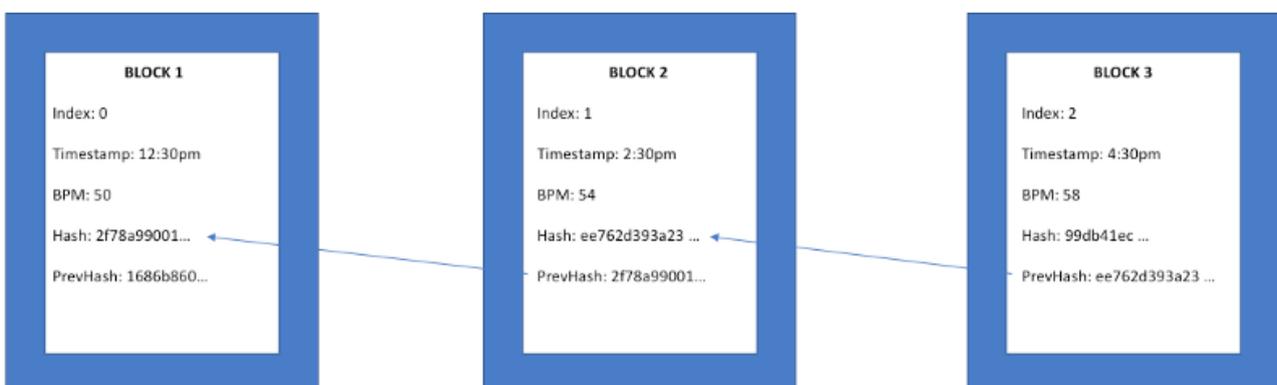


Figure 1. Simple Blockchain Model (Coral Health 2018)

<sup>1</sup> Hashes are calculated with a hash function such as SHA and RIPEMD. The function takes a input string, performs a series of operations, and output another string of fixed length, the hash. (Madeira 2019)

Blockchain has several important characteristics. First of all, blockchain is immutable, which means once a block is added to chain, it cannot be changed. If a block is tampered, its hash will be different from the PrevHash stored in the next block. Therefore, no one can secretly change the data stored on the chain. Moreover, it doesn't allow single point of authority, which means no single party in the network has complete control over the data stored on the blockchain. Before a new block is added to the chain, an algorithm checks that the block satisfies all agreed-upon features, and this also ensures that all parties on the network has the same chain (Zheng 2017). This makes blockchain a decentralized technology, and is extremely useful for increasing data transparency.

To store data with blockchain, there are two options: on-chain and off-chain. Since blockchain was originally designed to store bitcoin transactions, its protocols or big transaction fees limits that only a small amount of data can be stored on-chain, usually in range of kilobytes or less with one kilobytes equals to approximately 500 words (Marx 2018). Therefore, a reasonable solution is storing the hash of the data on-chain, while storing the actual files and the corresponding block hashes (TX-ID) off-chain as shown in Figure 2.

There are two common options for storing data off-chain. The first one is traditional database or cloud storage. However, there are several problems with this first option. Once the files are uploaded to cloud or inserted into database, they are once again controlled by one central point of authority, such as Google, Microsoft, Oracle and so on. Not only transparency could be lost, but also if the company decided to close down the storage service, data itself could be lost as well (Marx 2018). This leads to the second option - decentralized storage. In decentralized storage, data is distributed across many nodes on the network, and files are broken apart and stored on various nodes. So no single node has the entire file and breakdown with one node will not affect the others, so files are at a much lower chance of being lost permanently (Marx 2018). One such project is the InterPlanetary File System (IPFS).

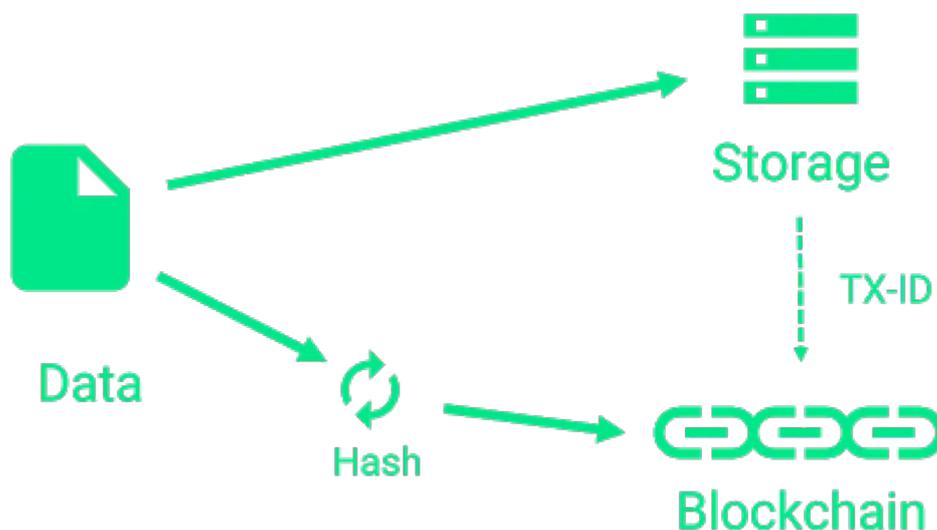


Figure 2. Storing data with Blockchain (Marx 2018)

### **Section 3.2 IPFS**

Just like the Hyper Text Transfer Protocol (HTTP) which the internet is based on today, IPFS is an internet protocol. However, unlike the location addressed HTTP where users get information from central servers according to the IP address, IPFS uses content addressing and a peer-to-peer (P2P) network in which users can share files directly with others in the network (Curran 2018). This is illustrated in Figure 3.

This gives IPFS several advantages. Since HTTP is location based addressing, if the server is down or the webpages are deleted, the files are not available anymore and useful information could be lost

(FortKnoxster 2017). Also HTTP is centralized, so access to data can be slow depending on where the server is, or even restricted, such as Google, Youtube, Facebook and so on are all blocked by the government of China (Carson 2015).

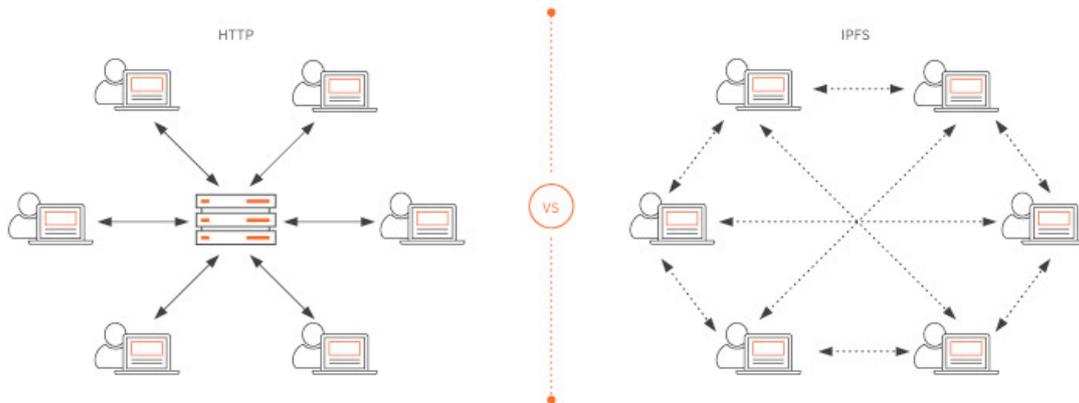


Figure 3. HTTP vs. IPFS (Curran 2018)

On the other hand, IPFS is content based addressing, meaning that the link of each file is composed of its unique hash, and every node in the network can choose to keep the files it is interested in. Each file is broken into multiple IPFS objects and linked together by an empty object as illustrated in Figure 4 (Fanzil 2019).

This makes sharing and downloading files much faster, since users not only can get data from the closest node which has a copy of it, but also can download parts of a file from different nodes at the same time instead of downloading the entire file from a single server (Curran 2018). Moreover, since IPFS is decentralized, all files on the network is publicly visible and cannot be blocked. Thus, transparency is preserved. A real life example happened in Turkey, 2017. Turkish authorities blocked access to Wikipedia throughout Turkey, but activists created a copy of Wikipedia on IPFS and made it available again (Dale 2017).

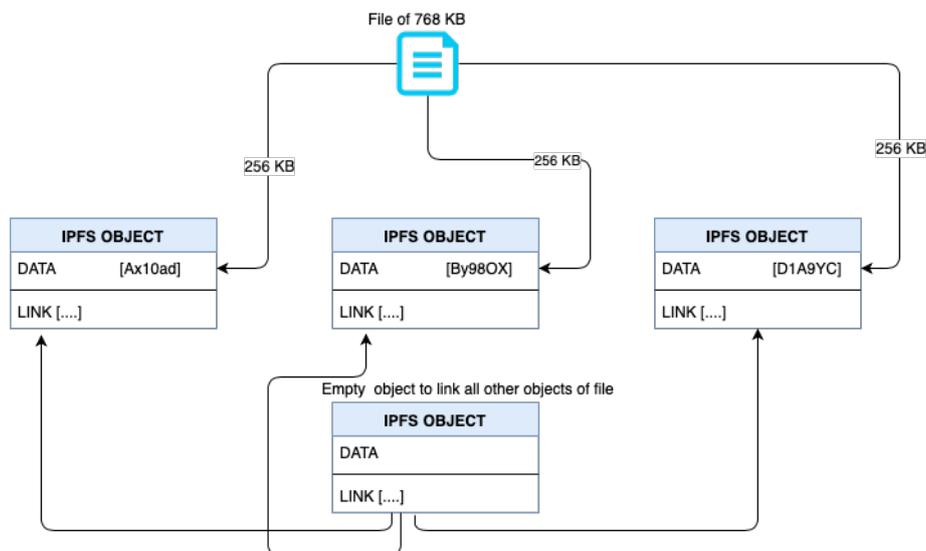


Figure 4. IPFS model (Fazil 2019)

**Section 4. Discussion of the Model and Open Data**

Our proposed model is to use blockchain together with IPFS to create a completely decentralized application that holds important college data. The issue with publishing data while worrying that

private information can get in the wrong hands can be absolved using the model we proposed. All the files will be stored on IPFS, and the immutable, permanent IPFS links will be placed into blockchain as shown in Figure 5 (FortKnoxster 2017).

Since IPFS tracks version history, using blockchains and IPFS can ensure that annual data will be preserved permanently on the network. In addition, every modification to the data will be visible to the public. Since the consensus algorithm will check the information to be added on the chain, sensitive information is highly unlikely to be published or accessed by parties outside of the network. Therefore, IPFS makes a good candidate for our purpose -- increase transparency of college data.

This model can benefit three parties involved in higher education: the students, universities/colleges, and the government.



Figure 5. Use of IPFS with Blockchain (Coral Health 2018)

Prospective students can see previous versions of the college annual data and discover differences, improvements or setbacks on the network. Thus, having access to more holistic information, these students can make well informed decisions about their future. Universities and the government will have access to organized and clearly presented data. This will make it easier for them to analyze trends, discover issues, and fix problems within higher education. Moreover, governments can determine which universities and colleges will be a positive investment for placing their government aid. An effective data system will lead to effective decision making, for all parties involved.

## Section 5. Conclusion

Federal organizations, in general, need to promote the existence and importance of the availability of data. There is no use in making data more accessible to the public without active citizen engagement, because only involvement can push for development. An ideal future next step is to implement this data network through a decentralized application built with blockchain and IPFS. When building this application, we can learn from research or existing implementations of blockchains or IPFS in different fields. Since the use of this decentralized data system is flexible, we strongly encourage governments to store other federal data on the network, which would increase comparability of data and minimize the time in finding problems.

## Acknowledgements

We would like to thank our close friends and family for the support they have given us throughout the year.

## References

- Association of American Colleges & Universities. n.d. “Indicators of Equity in Higher Education in the United States.” Accessed May 4, 2019. <https://www.aacu.org/aacu-news/newsletter/indicators-equity-higher-education-united-states>.
- Berg, E.A. and Hanson, M. 2018. “Putting the ‘Evidence’ in Evidence-Based: Utilizing Institutional Research to Drive Gateway-Course Reform” in *Improving Teaching, Learning, Equity, and Success in Gateway Courses: New Directions for Higher Education*, Number 180. Koch, A.K. (Ed.). John Wiley & Sons.
- Carson, Biz. 2015. “9 Incredibly Popular Websites that are Still Blocked in China.” *Business Insider*. Accessed April 10, 2019. <https://www.businessinsider.com/websites-blocked-in-china-2015-7>.
- Coral Health. 2018. “Learn to securely share files on the blockchain with IPFS!” *Medium*. Accessed April 2, 2019. <https://medium.com/@mycoralhealth/learn-to-securely-share-files-on-the-blockchain-with-ipfs-219ee47df54c>.
- Curran, Brian. 2018. “What is Interplanetary File System IPFS? Complete Beginner’s Guide.” *Blockonomi*. Accessed April 2, 2019. <https://blockonomi.com/interplanetary-file-system/>
- Dale, Brady. 2017. “Turkey Can’t Block This Copy of Wikipedia.” *Observer*. Accessed April 2, 2019. <https://observer.com/2017/05/turkey-wikipedia-ipfs/>
- Esters, Lorenzo L. “Making an Impact in Higher Education Equity.” *Strada Education Network*. Accessed May 4, 2019. <http://www.stradaeducation.org/making-an-impact-in-higher-education-equity/>.
- Fazil, Usman. 2019. “IPFS: A Distributed File Store.” *Block360*. Accessed April 2, 2019. <https://block360.io/ipfs-a-distributed-file-store/>.
- FortKnoxster. 2017. “How the IPFS Concept Can Change the Internet and the Storage Distribution.” *Medium*. Accessed April 2, 2019. <https://medium.com/fortknoxster/how-the-ipfs-concept-can-change-the-internet-and-the-storage-distribution-c6c13283f12d>.
- Gandhi, Rohith. 2018. “InterPlanetary File System(IPFS)—Future of the Web.” *Medium*. Accessed April 2, 2019. <https://medium.com/coinmonks/interplanetary-file-system-ipfs-future-of-the-web-c45c955e384c>
- National Center for Education Statistics (NCES). Integrated Postsecondary Education Data System. Accessed April 10, 2019. <https://nces.ed.gov/ipeds/use-the-data>.
- Koch, A.K. (Ed.). 2018. *Improving Teaching, Learning, Equity, and Success in Gateway Courses: New Directions for Higher Education*, Number 180. John Wiley & Sons.
- Madeira, Antonio. 2019. “How Does a Hashing Algorithm Work.” *CryptoCompare*. Accessed April 10, 2019. <https://www.cryptocompare.com/coins/guides/how-does-a-hashing-algorithm-work/>.
- Marx, Lukas. 2018. “Storing Data on the Blockchain: The Developers Guide.” *Malcoded*. Accessed April 2, 2019. <https://malcoded.com/posts/storing-data-blockchain>.
- Open Data Charter. 2015. “Principles.” Open Data Charter. Accessed July 21, 2018. <https://opendatacharter.net/principles/>
- Open Data Institute. 2017. “What is Open Data and Why Should We Care?” Accessed May 2, 2019. <https://theodi.org/article/what-is-open-data-and-why-should-we-care/>.
- Ortega, Jennifer. 2017. “Student Right to Know Before You Go Bill Introduced.” *EDUCAUSE*. Accessed May 4, 2019. <https://er.educause.edu/blogs/2017/12/student-right-to-know-before-you-go-bill-introduced>
- Schneider, M. 2017. Reforms to Increase Transparency in Higher Education: Testimony before the House Subcommittee on Higher Education.
- Smith, Ashley A, 2018. “States Attempt Closing Racial Gaps to Improve Graduation.” *Inside Higher Ed*. Accessed May 4, 2019. <https://www.insidehighered.com/news/2018/08/21/states-showing-some-progress-closing-racial-equity-gaps>.
- Student Data Principles. “The Principles.” Accessed April 15, 2019. <https://studentdataprinciples.org/the-principles/>
- US House. 115th Congress, 1st Session. *H.R. 4375, STEM Research and Education Effectiveness and Transparency Act*. Act. Washington, Government Printing Office, 2017. Passed in 2017.
- US House. 115th Congress, 1st Session. *H.R. 4479, Student Right to Know Before You Go Act of 2017*. Act. Washington, Government Printing Office, 2017. Introduced in 2017.
- US Senate. 115th Congress, 1st Session. *S. 1121, College Transparency Act*. Act. Washington, Government Printing Office, 2017. Introduced in 2017.
- West, M. S. 1998. The Historical Roots of Affirmative Action. *Berkeley La Raza Law Journal*, 10(2): 607.
- Zheng, Zibin. 2017. “An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends.” *6th IEEE International Congress on Big Data*. 10.1109/BigDataCongress.2017.85.