Blockchain Primer

Most simply put, blockchain is technology built on the concept of the distributed ledger. So, what does this actually mean? In a well-functioning blockchain, the original moment of data creation is recorded in the blockchain ledger as the original "block."[1] This transaction and each subsequent transaction after this original entry updates the ledger. The ledger is replicated on all the nodes participating in the blockchain, forming a distributed ledger. Through this distributed recording mechanism, the blockchain becomes immutable and blocks can be traced back to the original entry and every other related entry in that same lineage. An apt analogy to how blockchain works and how it can transform current technology and systems is by comparing it to genealogy and the concept of the family tree. Currently, any genealogist trying to reconstruct a family history has to rely on what is known about the family and do research to reconstruct familial links by visiting census data, property records, immigration records, and so on. This is a long and laborious process depending on the level of data the genealogist desires and is able to acquire. When that family tree has been developed, it can be compared and connected to the research of other genealogists on any of the popular genealogy sites. The family tree can then be compared to other family trees for overlaps and validation. If blockchain were used as the underlying technology, then every individual in the verified family tree could be established as one entry on the blockchain, created out of a "transaction" from two previous blocks. Thus, each record is linked to its preceding records and, by default, to every future record. Blocks within a genealogical blockchain could have data encoded to provide additional information on the individuals such as names, date and place of birth, height, eye color, agencies involved in adoption, links to genetic services, and so on. Thus, the blockchain could provide a verified record of the entire family tree at the press of a button in perpetuity.

Of course, this is an oversimplified representation of a blockchain. In the world of blockchain, this kind of diagram is called a Merkle tree. The original paper that introduced the Merkle tree was published in 1980 and established the basics of a blockchain protocol and how it could be cryptographically secured.[2] The Merkle tree logic allows for a very sophisticated and trusted algorithm to create unique identifiers for each block. This unique ID, called a "hash" in blockchain language, is a major feature in securing the blockchain and creating the immutable records that confirm the integrity of the data that is stored. The logic underlying the Merkle tree reduces the computational power required to verify the integrity of the blockchain because of the method by which hashes are created and linked to the preceding block in the blockchain. This mechanism affords participants in the blockchain a very high degree of security and privacy and has led to the Merkle tree becoming the basis of blockchain. Figure 2.1 is an example of a Merkle tree. The "uberblock" at the top represents the

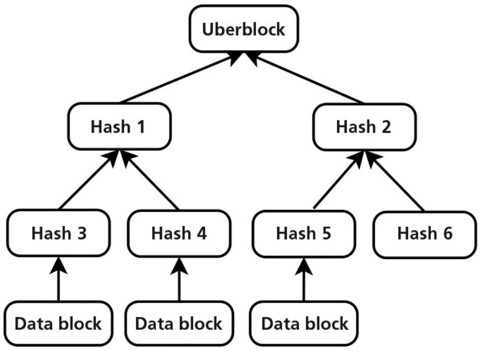

Figure 2.1
Example of a Merkle tree, a simplified representation of a blockchain structure. Source: "File:ZFS Merkle Tree 2.svg" by Markus Then is licensed under CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0/), https://commons.wikimedia.org/wiki/File:ZFS_Merkle_Tree_2.svg. Figure has been adapted; the bottom row has been updated from "Datenblock" to "Data block" and "Hash 6" has been added.

original block and each block below relates back to the original block.

However, what makes the blockchain so powerful as an application is its basis in distributed computing. In other words, the verification of the current block entry against all of its predecessors requires the availability of a network of computers that run the background checks required to validate the current record. In the most secure applications, the background checks have to be confirmed by a majority of participants in the blockchain (although a lower threshold can also be specified). The most significant benefit of this approach is that hacking the network becomes close to impossible, and even if it could be done, it would be very expensive. If one were to try to alter or forge a record, then the entire lineage of the record back to the original kernel would have to be changed, which in a distributed ledger would require knowing and hacking all of the computing nodes involved in retrieving and verifying those records. This introduces a significant degree of complexity that, although not impenetrable (in theory), does present a significant challenge that may outweigh the motivations of potential hackers. Consequently, blockchain has been considered in a wide variety of applications and scenarios where security and immutability are significant concerns.

The distributed ledger combines the technology underlying distributed computing with the concept of the ledger used in accounting. The distributed ledger is a digitized version of the paper ledger where transactions are recorded as they occur, thus providing the accounting and documentation required to ensure that transactions have taken place. In a ledger, one might record amounts, parties involved, time of transaction, and other pertinent information. The distributed ledger takes this information, places it online, and distributes identical copies of the ledger to all the computers in the system, thereby ensuring that validated copies exist in multiple places (similar to the LOCKSS principle). When a new entry (or in blockchain language, a new block) is added to the ledger, the distributed ledger is automatically updated across all nodes in the system. Every subsequent transaction is now added to this new and updated ledger.

Thus far, the major application of blockchain has been for cryptocurrency. Cryptocurrencies, such as Bitcoin, Ethereum, Ripple, and many others have been developed on the concept of the distributed ledger. They have come into existence to act as alternatives to traditional government-backed currencies such as the dollar or euro. As alternative quasi-currencies, cryptocurrencies have leveraged blockchain and enabled an entirely new global network of currency transactions. This has been done by employing blockchain as a distributed database working on computers located all over the world. These computers work on a system of randomly assigned verifications needed to maintain the integrity of the blockchain. Computers participating in this verification process are called “miners.” They engage in calculations of varying intensity and difficulty that work on solving the unique codes used in every block. This unique code or hash is a cryptographically-secured string of numbers. As an incentive for participating in this process, at certain intervals, miners are issued a reward. Typically, miners receive a unit of value of the blockchain’s currency seeking validation (e.g., one Bitcoin). This process is important, as we will explore later, since there has to be incentive in the system to ensure that enough computers are participating in the verification, which in turn ensures the security and speed of the blockchain. To put this incentive into perspective, one Bitcoin at its peak value in 2017 was worth over $20,000. Although the price of Bitcoin fluctuates significantly and in 2019 Bitcoin has been trading in a range from the low $3,000s to just over $13,000, the incentive often outweighs the costs of time and energy expended in the process of mining.[3] To further explain the principle behind how Bitcoins are awarded, imagine 100 miners are working on the calculation tasks required to verify the blockchain. The blockchain may have been set up to award a token after every 1,000th verification is completed. If miner 1 solves verification number 999, it would get nothing. If miner 2 solves verification number 1,000, it would receive one Bitcoin. If miner 3 solves verification number 1,001, it would get nothing, and so on, until somebody solves verification number 2,000. Thus the process provides both incentive and motivation, but also a sufficient amount of randomness so that all are engaged to the best of their abilities. In libraries, the financially motivated incentive mechanisms of cryptocurrency do not exist. However, other incentives may have to be developed. Depending on the applications, consortial agreements may predetermine contributions from those participating in the blockchain. For example, suppose thirty libraries decide to develop a blockchain and contractually dedicate a certain amount of computing power to allow for the blockchain to always be available and up to date. In that case, the contract is in place; other incentives are not needed. However, larger public blockchain applications would need new incentive models that would appeal to those required to participate. What exactly those will be will depend on the application and who is expected or required to participate.

Private versus Public Blockchains

Blockchains have two main variations that have significant impact and influence over how they function, who can participate, and who has control over them. A blockchain can be private or public.[4] Private blockchains are, as suggested by the name, exclusive in nature. Only those invited and authorized can participate in a private blockchain. This creates a controlled environment with a limited number of authorized participants. Public blockchains are exactly the opposite. They are open networks that anyone can participate in, adding and verifying transactions. Unlike private blockchains, public blockchains are typically decentralized. The network protects itself through scale and enabling any member of the blockchain to audit and validate the data. Typically, this kind of blockchain is involved when discussing cryptocurrency applications. Conversely, in a private blockchain, the distributed network is limited, and all users are known. Whether a blockchain is private or public is up to the developer of the application. This has to be determined at the very beginning stages. In libraries, we may find that both types of blockchains have applications and can be employed depending on the problem at hand. The thought starters provided in the next chapter will address the benefits and challenges associated with these two types of blockchains and will present use cases that consider the benefits of private versus public blockchains.

Power Consumption and Computing Power

Public blockchains are designed to provide immutable records of transactions. The underlying value of a public blockchain is derived from trust established by the decentralized system ensuring that single actors and coordinated schemes to subvert it will be unsuccessful. The blockchain is available to all and can be verified by any member. For future transactions to be validated, a majority of members need to verify the blockchain, also known as "proof of work." This process of verification is complex and needs to happen quickly in order for the blockchain to function efficiently. As the blockchain grows and transactions increase, the level of complexity grows, which leads to increased need for computing power. As a result, the increasing demands on computers and processors to continuously verify the blockchain lead to massive energy consumption. Various large blockchain applications have been estimated to consume more electricity than entire nations over comparable periods of time.[5] However, this issue will not typically arise in the way most libraries would employ blockchain technology. If a library were to deploy a public blockchain, then the transaction volume would not be even close to the transactions required by cryptocurrencies such as Bitcoin or Ethereum. The reason for this discrepancy is in the frequency and in the increments with which cryptocurrencies trade. Cryptocurrencies can trade in fractions to the eighth decimal (i.e., 0.00000001 Bitcoin). Thus, a Bitcoin can be divided and subdivided and recombined over and over. This complexity, combined with the frequency at which currencies can be exchanged, is far in excess of any transactions that are likely to occur in libraries, such as circulation data or patron data. (Such transactions also by their nature are unlikely to be divided down to the eighth decimal.) Furthermore, the "proof of work" requirement could be set at a different and lower rate from what is required in cryptocurrency applications, thus significantly reducing the need for computing power. In a private blockchain, the permissions for those who participate can be set very differently from a public blockchain—so differently in fact that power consumption could be much better managed because the private blockchain with all participants known would be a trusted and reliable recording mechanism. Proofs or verifications might not be required, and certainly not in the same way that a public blockchain would require. Furthermore, a private blockchain would serve well in many library applications since it would allow for faster verification. It would not need the 51 percent proof of work consensus mechanism required to prevent fraudulent activity in many public blockchains due to the anonymity of the users and miners. There are many versions of private blockchains emerging, and they are being branded in a number of ways. For example, Hyperledger was developed by the Linux Foundation in 2016 and has found significant support from many commercial entities across the spectrum of consulting, banking, and manufacturing industries.

Hyperledger
https://www.hyperledger.org

What Can Be Encoded in a Block?

Blocks on the blockchain are information containers. They can hold a wide variety of content. At the very least, a block stores its own unique identifier, or "hash," that links it to all blocks preceding it and all subsequent blocks. Each block is uniquely identified through its hash, which is automatically generated and ensures there is no ambiguity between different blocks. However, much more can be stored in a block. In addition to this identifier information, blocks can store data of all kinds related to a transaction, such as the following:

- time
- date
- measurements (e.g., height, width, weight, etc.)
- text

- transactional information
- computer code that can trigger actions (usually referred to as “smart contracts”)

These various kinds of data can be automatically generated or can be manually added. In practice, the data stored is readable across blocks. The data in the blocks can then be queried and analyzed. The size of what can be encoded in a block is limited only by the specifications set by the creator of the blockchain. The blocks can be small and allow only a few kilobytes of data, or they can be quite large and allow several megabytes of data. As the technology evolves, it will be possible to attach PDF and image files, audio files, video files, and files of other formats that have not been previously associated with blockchain. One limitation thus far has been related to the computing power required to process building, storing, and verifying blocks. Since every block includes information linked to previous blocks, there has been some concern about the overall size of the blockchain database. As technology advances and computing power increases, the boundaries of these limitations will be tested and will expand. It remains to be seen whether Moore’s Law,[6] the increasing speed of the internet (some will remember the early days of the public internet and dial-up modems), or the evolution of the cell phone to the current smartphone is an apt analogy. However, as with all successful technologies, increasing adoption will lead to increasing investment, and ingenuity in how the technology can be optimized and improved will follow.

Blockchain and Privacy

One of the keys to blockchain technology that has made it viable for cryptocurrencies is the privacy features. These features constitute a key component of the blockchain and could be considered built in by design. There are three main areas in which the blockchain is particularly strong:

- public and private keys
- the public blockchain
- the private blockchain

Public and Private Keys

Participants in a blockchain have to gain access to it. In order to do so, a participant has to register or be issued a private key. Depending on the blockchain rules, the participant’s public key can be issued by the owner of the blockchain, or the participant can autogenerate a key, which ensures even greater privacy. The public key is a complex alphanumeric sequence that is unique to the participant. However, participants in the blockchain are not limited to only one private key and thus can have multiple accounts. A complex algorithm converts private keys to public keys. Public keys are used for the record-keeping of the blockchain. The algorithm used to derive public keys from private keys cannot be reverse engineered, which ensures that the private key always remains private. In the blockchain, when a transaction is initiated, the public key is recorded with the blocks to provide accountability of the transacting party. The public address can be queried, and transactions can be traced back to the public key. Because public keys cannot be reverse engineered to the private key, owners of a private key remain anonymous unless they reveal their private keys. Due to this extreme privacy function, private keys cannot be recovered once lost. It is worth noting that the data that has been encoded in the blockchain remains there forever due to the immutability of the blockchain. There are countless stories in cryptocurrency of lost private keys, which means that the coins associated with those keys cannot be recovered by the original owner and thus are lost forever. This is akin to losing the keys to a treasure chest that has been hidden somewhere. In other words, the contents of the treasure chest still exist, as does the record of their existence. However, the contents have now become irretrievably lost. As it happens, many private keys have been lost. Estimates point to roughly 17 to 23 percent of all Bitcoins ever mined having been lost.[7] This can prove to be a challenge if a user were to lose their private key. However, this tradeoff in convenience has to be accepted if this level of privacy is desired. Unlike with a password to an email account, where a forgotten password can be retrieved by answering a few security questions to access the account again, a forgotten private key is irretrievably lost and the account is no longer accessible.

Public Blockchain

In a public blockchain, everybody can join. Using a private key, that has been converted to a public key allows anybody with an internet connection and a computing device able to run the blockchain software to participate. Since the public blockchain is a distributed ledger of all transactions, no single user can corrupt the data. When a transaction takes place on the blockchain, a new block is created. However, the block does not get added to the blockchain until it has been verified by a majority of the participants. Depending on the number of participants and a few other factors, this verification can take place in real time or may take a longer amount of time. Most importantly, though, the consensus required to verify the blockchain ensures that the security, privacy, and integrity of the blockchain are maintained.

Private Blockchain

In a private blockchain, the owner of the blockchain has significant influence over its design and subsequent operations. As a result, a private blockchain is a less secure and private type of blockchain. Here, the participants in the blockchain are known, and blocks can be altered at the owner's discretion. While this may pose a privacy challenge, it does not mean that the blockchain cannot be maintained with strict privacy controls in place. Therefore, depending on the desired application, a private blockchain could be a viable application that ensures security.

How Do Blocks Talk to Each Other?

The blocks in a blockchain talk to each other by being linked in a very linear way. The Merkle tree diagram (see figure 2.1) provides a visual representation of how each block is linked back to the preceding block. In blockchain, the hash from the first block is combined with the hash from the second block to make a new, unique combination. The next block combines the earlier block and the new information by adding its unique signature. The hash itself is an encrypted and complex alphanumeric combination, ensuring that the combination is unique. Here is a very simplistic way to think of the way hashes work:

> Original hash: A1
> Next block hash: A1 + new hash, i.e., A2 → new block hash of A1A2
> Next block hash: A1A2 + new hash, i.e., A3 → new block hash of A2A3
> And so on . . .

Through this combination of hashes, the entire blockchain can be verified and traced back to the original block. The ingenuity of this method is that while one can always trace known hashes backward, one cannot predict future hashes. If A1A2 is known, then hackers could go back and try to alter A1. However, they would have to alter all the preceding blocks in the blockchain. If A2A3 has already been created, the blockchain would detect the fraud attempt. More importantly, since the blockchain lives in the cloud and is replicated on many computers, there will always be copies of the original blockchain available to verify against. In cryptocurrency, where this kind of attack would be a grave concern, the developer community has decided on a concept called "proof of work." Proof of work requires 51 percent of the network to confirm the transaction, thereby making a coordinated attack on the blockchain nearly impossible. This built-in security ensures, as the blockchain community refers to it, immutability—that is, that the block cannot be altered after it has been created, verified and added to the blockchain.

Problem or Solution: Which One Came First?

In libraries we deal with myriad challenges on a regular basis. We try to create engaging environments. We try to work within our budgets. We work with our patrons, users, scholars, students, clients, or whatever other user-specific term is employed in your organization. We work with each other across divisions, different locations, consortia, and so on. We try to measure and share the value we add to our environments. All of these challenges are looking for solutions. However, as the old aphorism reminds us, "If your only tool is a hammer, all problems look like nails." Thus the question arises, "What problems are we able to solve with blockchain?" Throughout this report, we will think through the "why" and "so what" related to blockchain as a solution to the problems and opportunities presented. For what it is worth, libraries function, and function well at that. We share catalogs and records. We have patron records in our databases. We manage our collection budgets. We issue library cards. And, to say the very least, we are keenly aware of issues related to privacy. The author of this report posits that some of these areas could be significantly improved by employing blockchain as a technology. However, even though blockchain technology can address these issues and concerns, often the implementation will raise new issues. Ultimately, each case will have its own specific context that will decide whether the technology is transformative and of sufficient value for consideration and implementation in your organization. In chapter 3, we will present thought starters so that you, the reader, may consider the various opportunities and challenges to make your own determination about whether blockchain is an appropriate solution to your problems and whether it meets the ethical standards you hold yourself to.

Why Should Libraries Care about Blockchain?

So what? We have now established some of the basics of blockchain, but why should libraries care? Libraries should care because blockchain is here to stay. Many corporations have bought into the idea of blockchain to support their enterprises. As acceptance grows and use cases emerge, our library community will be presented with applications based on blockchain technology. It is not farfetched to think that library systems will be developed leveraging blockchain. Perhaps our next-generation integrated library systems will be built on open standards and blockchain will be used to secure user records in the system. Thus, it behooves us as libraries to be informed and at least conversant on the topic of blockchain so that we can truly evaluate whether we are being presented with feasible applications and systems or just alluring trends and marketing pitches. Applications that we have not thought of yet will be developed that leverage blockchain. Therefore, we have a significant opportunity to contribute to the development of blockchain technology within libraries, museums, and archives. Some opportunities for the use of blockchain will be related to the scholarly record, research, funding mechanisms, and so on. Thus, it would be wise for libraries to prepare for these conversations. Another likely important connection will be linking blockchain with the emerging technologies of big data and artificial intelligence. However, perhaps the best answer to why libraries should care about blockchain is because the technology provides us with the possibility to develop significantly improved systems as compared to where we are today.

In the thought starters in the next chapter, we will explore some of these concepts and provide more details on how blockchain may be employed in various libraries-related scenarios.

Notes

1. For a fascinating account of how a cryptocurrency was launched, you can access Molly Webster, reporter, "The Ceremony," podcast, produced by Matt Kielty and Molly Webster, Radiolab, WNYC Studios, July 14, 2017, https://www.wnycstudios.org/story/ceremony.
2. Ralph C. Merkle, "Protocols for Public Key Cryptosystems," in *Proceedings of the 1980 IEEE Symposium on Security and Privacy, Oakland, California, USA, April 14–16, 1980* (Silver Spring, MD: IEEE Computer Society Press, 1980), 122–34, http://www.merkle.com/papers/Protocols.pdf.
3. CoinMarketCap, Bitcoin statistics, accessed September 7, 2019, https://coinmarketcap.com/currencies/bitcoin.
4. Demiro Massessi, "Public vs. Private Blockchain in a Nutshell," Medium, December 12, 2018, https://medium.com/coinmonks/public-vs-private-blockchain-in-a-nutshell-c9fe284fa39f.
5. Garrick Hileman and Michel Rauchs, "2017 Global Cryptocurrency Benchmarking Study," April 6, 2017, last revised September 20, 2017, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2965436; Jingming Li, Nianping Li, Jinqing Peng, Haijiao Cui, and Zhibin Wu, "Energy Consumption of Cryptocurrency Mining: A Study of Electricity Consumption in Mining Cryptocurrencies," *Energy* 168 (2019): 160–68, https://doi.org/10.1016/j.energy.2018.11.046.
6. "Moore's Law 40th Anniversary Press Kit," Intel, accessed September 7, 2019, https://www.intel.com/pressroom/kits/events/moores_law_40th/index.htm.
7. Louise Matsakis, "How WIRED Lost $100,000 in Bitcoin," *Wired*, May 28, 2018, https://www.wired.com/story/wired-lost-bitcoin; for an open source download on Bitcoin blockchain analysis, see Harry Kalodner, "BlockSci: A Platform for Blockchain Science and Exploration," Freedom to Tinker, Center for Information Technology Policy, Princeton University, September 11, 2017, https://freedom-to-tinker.com/2017/09/11/blocksci-a-platform-for-blockchain-science-and-exploration.