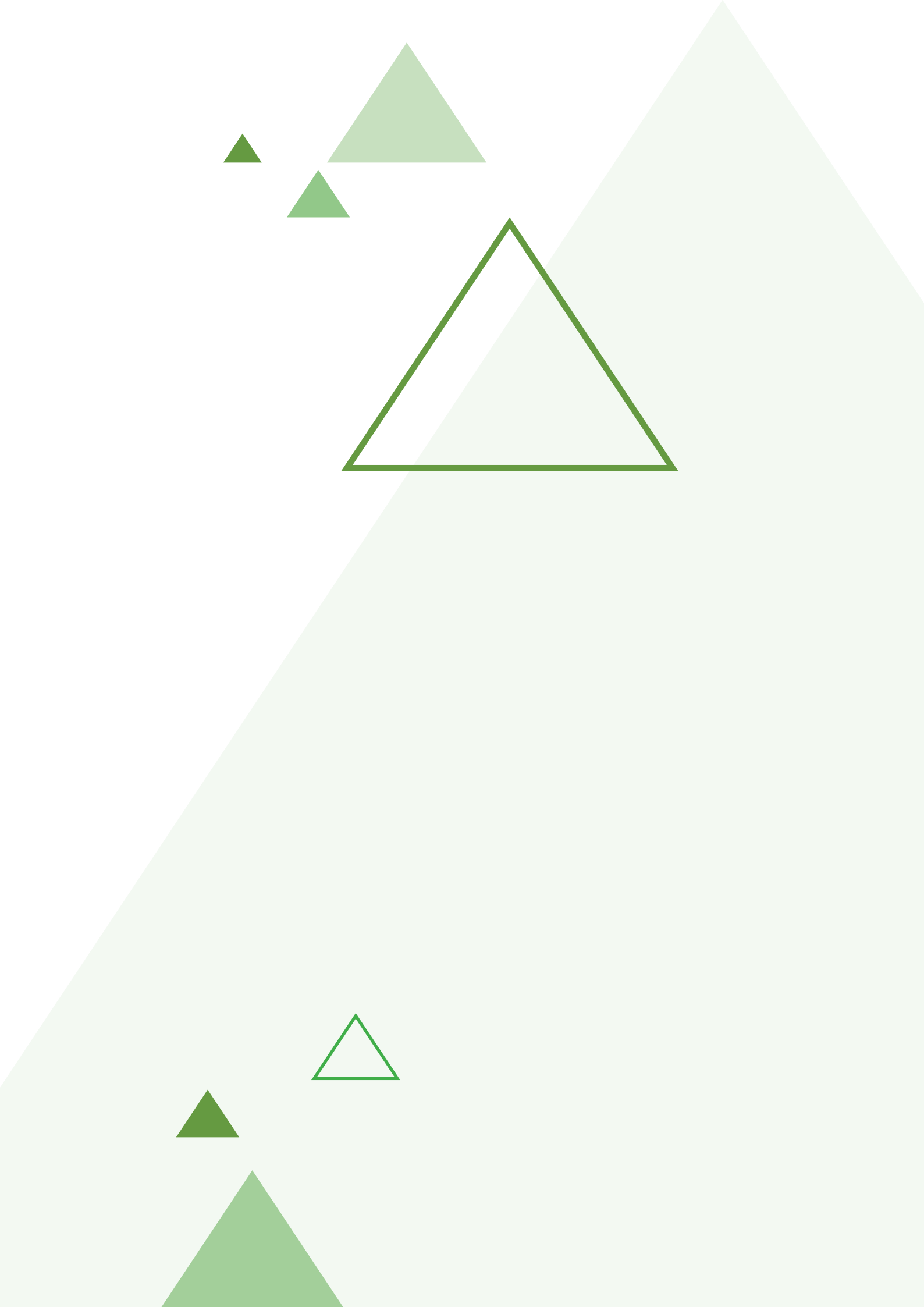
JUNE 2018

DIGITARY

Secure online credentials

Blockchain Position Paper

Andy Dowling
andy.dowling@digitary.net



Index

Overview

In 2018, it is difficult to have a technology conversation without somebody mentioning blockchain. While originally designed for the secure exchange of digital currency, this exciting and disruptive technology is being generalised and applied across a variety of problem domains. One area in particular is the issuing and verification of academic credentials.

For over 15 years, vendors and education providers around the world have successfully implemented solutions to enable the certification, exchange, and verification of digital academic records (see Appendix), yet blockchain is now being perceived as a perfect solution to digital credentialing.

This raises the question:

what does blockchain add to the digital credentialing space?

To explore this question, the Digitary team dived into the details of the blockchain architecture to understand how blockchain works and how it is being used to secure digital academic credentials.

We share our thoughts in this paper and outline our key observations. Our hope is that that this will progress the conversation around the application of blockchain to securing academic credentials.

Blockchain-based academic records

It is assumed that the reader has a rudimentary, high-level understanding of blockchain technology. It is important to qualify that there are different types of blockchain architectures:

- **Public** blockchain – openly-accessible networks such as Bitcoin/Ethereum;
- **Private** blockchain – where organisations run their own blockchains internally and control who can access them;
- **Permissioned blockchain** – usually shared blockchains run by consortia who collectively manage the blockchain and govern access.

In this paper, we are discussing **public blockchain only**.

We limit our scope to public blockchain because it is the most widely discussed solution for the secure certification and verification of academic credentials.

How is the blockchain approach different?

Solutions that are based on public blockchains are designed in a very different way to "traditional" digital credential architectures and exhibit the following main traits:

- **Decentralisation** – Public blockchains are based on the concept of an open and secure distributed ledger. There is no single point of access/truth/failure as many thousands of nodes around the world form the network and cryptographically verify transactions by majority consensus.
- **Cryptographic foundation** – There is no perimeter in a public blockchain and the system is open to all. The entire security of the public blockchain is based on cryptography. Advanced cryptographic techniques, combined with the sheer number of nodes in the network, make it extremely difficult, if not impossible to fabricate a transaction on the ledger.
- **Pseudonymous** – Public blockchains are designed for psuedonymity. While participants need to prove they own their cryptographic keys during transactions, blockchain cannot tell who they are.
- **Privacy and Transparency** – When certifying records on the blockchain, only the cryptographic hashes are stored "on-chain". The record itself is held privately by the learner "off-chain". This makes the blockchain paradigm good for the privacy of records. Also, blockchains publish a list of every action on their ledger and its respective output, to maintain transparency.

What does blockchain offer?

Architectures that use public blockchains implement a radically new approach to digital credentialing that promotes four key features:

- **Self-sovereignty** – giving learners "ownership", or full control over how and where their personal data is stored, while enabling them to prove ownership of those records;
- **Disintermediation** – removing the need for the record issuer (i.e. the education provider) to facilitate access to, or maintain academic records;
- **Trust** – through a robust and secure technical infrastructure that enables the issue and verification of academic achievements;
- **Immutability** – allow achievements to be captured, written and stored, digitally and permanently, without the possibility of modification.

We now examine each of these features in more detail and share our observations.

Self-Sovereignty

The ability for learners to access and share their digital records with third parties, without the intervention of the issuer, is something that some digital credentialling platforms have offered for over a decade.

Blockchain, however, takes things a step further by introducing the idea of "self-sovereignty", where:

- **the learner has *possession* of their record** (i.e. they keep the actual PDF, PNG, blockcert, etc.);

  and

- **the learner has the ability to prove that the record was issued to them, using their unique cryptographic key**.

These properties are central to the concept of "learner-owned" credentials and is what sets public blockchain apart from other architectures. Self-sovereignty provides two distinct advantages:

- **Learners do not depend on the systems of the issuer (or a vendor who serves the issuer), to access their information or prove its authenticity to others;**
- **The learner's record remains in the possession of the learner, and not in a repository that could potentially be subject to a data breach.**

Self-sovereignty places the learner at the heart of the digital credential ecosystem. However, for this to work, the learner must:

- **Keep their records safe, *for the lifetime of the record* or they will lose access to it;**
- **Keep their cryptographic keys safe *otherwise they cannot prove that they own the record*.**

Essentially, learners need to understand that their academic records are irrecoverable if lost, just like digital cash. This is even more difficult because academic credentials are accessed far less frequently than cash, so the likelihood of the learner forgetting their wallet passphrase is far higher over time.

All told, it seems unrealistic to expect a learner to keep their academic records and cryptographic keys safe for their career lifetimes. It is more likely that they will at some point need to go back to the issuer for a replacement.

We offer some observations:

> **#1: Learners do not need to "own" their record in order to share it with others, they only need access to it.**
>
> **#2: While self-sovereignty is empowering for the learner, it puts a burden on the learner to keep their records and keys safe and secure, for life.**

To make self-sovereignty work, learners (or, in the generalised case, individuals) need a standardised, mature, and widely-adopted method to enable them to store, backup, restore, and migrate their essential digital assets and identities, for life. This is not a trivial undertaking and while there is work ongoing in this space, it needs more time to mature before it will be generally adopted.

Not only this, but it will take time for individuals to become accustomed to the idea of being custodians of their own digital records, and that these records are irrecoverable if lost.

Disintermediation

Disintermediation takes records out of issuer repositories and put them in the hands of the learners, with independent verification via the blockchain which doesn't depend on the issuer. The goal here is to mitigate against the risks of data loss at the issuer's end as well as achieving learner self-sovereignty.

Part of the reasoning here is that data repositories can be targets for hacking and if compromised, the likely outcome is a large-scale data breach. Also, if an issuer's repository disappears (war, natural disaster, issuer goes out of business, etc.), the affected learners can no longer prove their achievements.

For disintermediation to work in practice, an issuer must never need to re-issue a record. We question if this is a realistic assumption, because even if an education provider certifies records on the public blockchain, they will still need a copy of each record issued, to cover the following scenarios:

- **The learner loses their record and it needs to be replaced;**
- **The learner loses their blockchain wallet keys and the document needs to be re-issued to a new set of wallet keys;**
- **The record needs to be re-written onto the blockchain because blockchain security becomes compromised;**
- **The issuer needs to revoke the record after it is issued;**
- **The issuer needs to maintain a record of what they issued for legal or compliance reasons.**

Even when certifying records on a blockchain, issuers still need to keep a copy of the records they have issued. Granted, there are risks with managing data repositories, but these risks can be managed with comprehensive information security and management practices. This is no different to what education providers do each day as their information systems interface with thousands of students online.

Observations:

> **#3: Blockchain does not remove the need for an issuer to maintain a repository of records.**
>
> **#4: Blockchain can provide the learner with a secondary verification method in the event that the issuers repository becomes unavailable, provided that the learner keeps their copy of their record.**

A learner-held copy can provide data continuity where the issuer's repository becomes unavailable. However, since the issuer needs to keep a copy of the record anyway, and the issuer is ultimately the source of truth for that record, it follows that the issuer's copy should be regarded as the authoritative record for as long as it is available at the issuer.

Trust

"Back to source" verification, involving a connection back to the issuer, is a simple, proven, trustworthy, and effective form of verification that relies on the actual record as maintained by the issuer. Verifiers only need to ensure that they are talking to the source of truth (the credential issuer), or a service that represents them (and this fact is published by the issuer), and it can perform all necessary checks for that record. Ensuring one is securely connected to a known website is a well understood procedure, as it has been the foundation for digital commerce for over 25 years.

Blockchain, for all of its underlying security and transparency mechanics, actually complicates verification because it places a new, complex, and relatively unknown technology as an intermediary between the verifying party and the issuer. The verifying party now needs to understand:

- **What the blockchain can and cannot verify;**
- **Whether or not they have a trusted app / connection to that blockchain;**
- **Whether the software / app they are using is comprehensively verifying a record.**

An important point here is that public blockchains are, by design, limited in what they can actually verify. While blockchains can verify that a record was issued at a point in time, they are designed for pseudonymity and privacy and so cannot identify the issuer of that record. While this is fine for cryptocurrency transactions, it is critical to verifying a degree that the issuer is the education provider they claim to be.

This is why solutions that use the public blockchain rely on "traditional" web-based services at the issuers website - to establish trust via issuer and comprehensively verify a record. Ironically, this contradicts the goal of disintermediation because the issuer is now involved in the verification process. This raises an awkward question: *if a blockchain doesn't offer complete verification, and you have to use the issuer's website to verify a record, then what value is blockchain adding here?*

Observations:

> **#5: By design, blockchain cannot tell if a credential issuer is who they claim to be;**
>
> **#6: Blockchain verification is actually more complex and less complete than more established online methods.**

Removing the need to use the issuer's website as a trust anchor will take time and depends on the ongoing work around distributed identities for blockchain (W3C DID[1]) and Distributed PKI (DPKI). While this is interesting work with potential, it will take time to understand whether it is fit for purpose in , and for it to become fully standardised, accepted, and to mature in production environments.

[1] See https://w3c-ccg.github.io/did-spec/

Immutability

Cryptography is a constant race between those making unbreakable codes and those trying to break them. Crucially, a public blockchain relies 100% on cryptography for its security. While the risk of broken cryptography is very low in the short term, algorithms inevitably weaken over time as a result of advances in cryptanalysis, mathematics, and computing.

Academic records need to remain authentic and unmodifiable - or *immutable* for life. This is hugely significant. While blockchain appears to offer immutability for numeric cryptocurrency transactions, this doesn't necessarily translate to certifying sensitive records. This is because only the cryptographic hash of the record is written to the blockchain, while the record itself stays "off chain" for privacy. This separation is where the long-term weak link lies.

If a learner wanted to modify their document, they need to engineer one with the same hash value as the original record (called a *hash collision*) and present the modified document in place of the original. If the hashes match, the fake document would be 100% cryptographically valid from a blockchain perspective. This type of attack is devastating because it can affect records retrospectively, is difficult to detect and the learner is not time-constrained in finding the collision.

Realistically, engineering a SHA-256[2] collision with today's technology is highly unlikely due to the sheer amount of computation required, however we cannot assume that no shortcuts will be found in the future.

The key point is this: when you write hashes of records to a public blockchain, **you are assuming that the hash algorithm will never be broken for the lifetime of that record** through advances in computing, cryptanalysis, and/or mathematics. This is a material, long-term risk and needs to be known and understood **before** writing any records to a public blockchain; because **if the algorithm is broken, all records hashed and written to a public blockchain with that algorithm will eventually be vulnerable to tampering.** This is effectively an information security timebomb that starts ticking from the time the record is written.

Secure hash algorithms have been broken in the past, including MD5 and SHA-1 (see https://shattered.io), and so it is only a matter of time for SHA-256. This is one of the reasons why we now have SHA-3 hash algorithms.

Observations:

> **#7: Blockchain does not necessarily guarantee the long-term immutability of hashed records.**
>
> **#8: Verifying a record "at source" with the issuer is safer than computing hashes and comparing them to hashes on a public blockchain.**

Mitigating against broken algorithms requires issuers to maintain a repository of records issued, so they can re-issue revised hashes of their records to the blockchain using a more secure hash algorithm in the future.

[2] SHA-256 is the hash algorithm currently used by Bitcoin and Ethereum

Considerations for issuers

Strategic cost

The decentralised nature of blockchain removes the learner's dependency on their education provider for access to and verification of their records. Therefore, education providers that use public blockchain for verification stand to lose key strategic benefits that traditional online platforms can provide:

- **Connectivity to their alumni, employers, and fellow institutions;**
- **An ability to collect aggregate statistics from the credential access and verification process (i.e. the mobility and career progression of their graduates) which can feed into their own strategic planning.**

Education providers are key to the adoption and population of digital credentials and they need to see both operational and strategic benefits when adopting any technology solution. An architecture that, by design, tries to separate education providers from their learners, when this doesn't need to be the case, doesn't seem to make sense from an issuer's perspective.

Observation:

> **#9: Disintermediation is strategically costly to education providers.**

Commercial cost

Public blockchains do not require the hosting of internal infrastructure, which is attractive in terms of cost of ownership for issuers. Instead, to write a transaction, the issuer must pay an amount in the relevant cryptocurrency, the actual cost of this in the issuer's local currency is only known at the time of the transaction.

While graduation/final documents such as degree certificates are relatively fixed in number for issuers, transcripts, which can be issued on a daily basis, will give rise to per-transaction charges that can fluctuate wildly. In December 2017, the Ethereum network transaction price was impacted when a new Blockchain application called "CryptoKitties" went viral (see http://www.bbc.com/news/technology-42237162).

The result was that the Ethereum blockchain became saturated with CryptoKitties traffic and transaction processing slowed significantly for several days. Due to the high load, the cost to write an Ethereum transaction jumped by an order of magnitude for a time.

Observation:

> **#10: The financial cost of writing a transaction to a public blockchain can be highly volatile.**

Conclusion

Blockchain is an exciting technology and is understandably generating a lot of interest lately. However, the application of any technology is just as important as the technology itself. Certifying academic records on a public blockchain is a light-touch solution to digital credential management that gets the blockchain and the learner to do the heavy lifting for the issuer.

This approach only works if learners never lose their keys or records, if issuers never need to keep a copy of their records, and if today's cryptographic algorithms remain indefinitely secure. These idealistic conditions don't hold in practice and this raises questions over the approach. Furthermore, we wonder why education providers would adopt a technology that, by design, aims to separate them from their learners unnecessarily.

#1: Learners do not need to "own" their record in order to share it with others, they only need access to it;

#2: While self-sovereignty is empowering for the learner, it puts a burden on the learner to keep their records and keys safe and secure, for life;

#3: Blockchain does not remove the need for an issuer to maintain a repository of records;

#4: Blockchain can provide the learner with a secondary verification method in the event that the issuers repository becomes unavailable, provided that the learner keeps their copy of their record;

#5: By design, blockchain cannot tell if a credential issuer is who they claim to be;

#6: Blockchain verification is actually more complex and less complete than more established online methods;

#7: Blockchain does not necessarily guarantee the long-term immutability of hashed records;

#8: Verifying a record "at source" with the issuer is safer than computing hashes and comparing them to hashes on a public blockchain;

#9: Disintermediation is strategically costly to education providers;

#10: The financial cost of writing a transaction to a public blockchain can be highly volatile.

Although we have identified issues with the public blockchain approach, clearly blockchain technology has significant potential. By leveraging blockchain technology in different ways, perhaps by using private or permissioned blockchains, and by allowing the technology to develop further, we might address some of the issues highlighted in this paper.

At Digitary, we keep an open mind and stand ready to embrace blockchain, only when it is ready and in a way that makes sense. Thankfully Digitary's platform architecture was designed with enough forethought to make such a transition simple for our millions of users.

In the meantime, we look forward to further exploring the blockchain conversation with our colleagues in the digital credentials and blockchain communities.

Appendix A
Proven Digital Academic Credentialing Systems

Digital credentialing platforms have existed for over a decade as a way to reduce credential fraud, increase efficiencies, and promote digital student mobility by replacing archaic paper workflows with secure digital alternatives. These approaches tackle digital credentialing in different ways and each approach has its own pros and cons:

	What it provides	How it works	Challenges
Central Repository	Allows employers and other third parties to check online whether a student or graduate has the qualifications they claim to have.	Participating schools publish records to a central repository which is then accessible to third parties who can interrogate the repository online.	Learners are not usually involved in the process, so privacy and consent can be issues that are often solved using manual, paper workflows. Security breaches are a risk.
Signed PDF	Education providers issue cryptographically signed PDFs can be shared by learners and independently verified.	Issuers generate digitally signed PDF files. Learners share the files with third parties. Third parties verify using PDF signature software.	Understanding of PDF signatures by third parties. Long-term validity of signatures is an issue.
Exchange Network	For education providers that do not trust student-submitted records, this enables education providers to send and receive academic records between each other.	Schools integrate with the network to send/receive digital records. They may also use a vendor to make this process easier.	Closed network. Some data protection / consent challenges remain as the student may not be involved in the exchange process. Additional components can be built to complement the network and obtain consent.
Hub and Spoke / Distributed Repositories	Enables education providers to issue their records. Provides learners with 24/7 access to their records and the ability to share their records with third parties without the issuer having to get involved in the process. Enables exchange with other regional providers by tying into existing networks.	Provides a fully integrated suite of systems that is a hybrid of (1) distributed repositories (one per school), connected via (2) an exchange network, with (3) student & third party portals to control access and (4) an external integration hub for international connectivity.	More suited to regional solutions and large scale projects (i.e. country-wide) as the underlying infrastructure and platform is heavily engineered to suit larger environments.

DIGITARY

Andy Dowling
andy.dowling@digitary.net


DIGITARY
Secure online credentials