Academic and Skills Credentialing Using Distributed Ledger Technology (DLT) and W3C Standards: Technology Assessment

Sthembile Mthethwa and Morne Pretorius
Council for Scientific and Industrial Research (CSIR)
Information and Cyber Security Centre (ICSC)
Pretoria, South Africa
smthethwa@csir.co.za, mpretorius2@csir.co.za

Abstract
The ongoing push for the 4th industrial revolution is setting the stage to digitise, persist and verify identity along with credentials. Academic and skills credentials are currently verified manually and have much scope for automation using cryptographic techniques but requires standardisation to facilitate future systems interoperability. The Distributed Ledger Technology (DLT) and World Wide Web Consortium (W3C) Verifiable Credentials (VC) standards presents the possibility to achieve this credential verification automation. To accomplish this, an understanding of various DLTs and requirements for a viable skills tracking system is important. Therefore, this research aims to access the selected DLTs against the assessment criterion presented and an analysis has been completed to determine which DLT is suitable for the proposed system. The DLTs are assessed in terms of their ability to support the rapid prototyping of such a system and provide recommendations to guide a future development path from the perspective of standards compliance. We conclude that few DLTs possess the maturity to provide proper requirements coverage due to the emergent nature of the DLT space. Additionally, this paper presents the high-level requirements to achieve a minimally viable solution that can demonstrate such digital credential verification in the academic and skills tracking context.

Keywords: distributed ledger technology, blockchain, standardisation, verifiable credentials, skills tracking

1. Introduction

The current system to verify qualifications in South Africa (SA) is known as the South African Qualifications Authority (SAQA) verification service, where all achieved qualifications are stored in the National Learners' Records Database (NLRD) (Ntshangase & Msosa, 2022). This is a centrally hosted database that will require some changes to improve the efficiency of their current verification process. A key challenge listed in their strategic plan for 2020/21 – 2024/25 is financial sustainability due to the current system having too many manual elements: *"Many of our processes are manual, outdated, and time-consuming. With adequate resources, SAQA can automate and streamline its processes; employ artificial intelligence to repetitive processes where necessary; improve productivity; and develop innovative solutions to complex problems"* (South African Qualifications Authority (SAQA), 2019).

From a security perspective, there are societal issues related to unethical human behaviour that technology will not solve (Schneier, 2019). For example, adopting the World Wide Web Consortium (W3C) Decentralised Identifier (DID) and Verifiable Credentials (VC) standards could give the current system the ability to identify, track and audit all system actors, but it cannot identify malicious actors that issue fraudulently signed credentials. There have been innovations in cryptographic threshold signatures that could improve the situation (Sonnino et al., 2018) and disincentivise unethical behaviour or impersonation by requiring a threshold number of actors from a quorum to jointly authorise academic credential issuers such as universities. This quorum of signatures could represent authoritative entities from across the world, or from within the country that acts as the Accreditation Authority (AA) (Gräther et al.,

2018) to prevent lone bad actors from compromising the system's integrity.

From an efficiency perspective, there could be much to gain when adopting the mentioned W3C standards and associated technologies to streamline administrative overhead as it becomes automated through cryptographic signature verification, and also fosters the move towards self-sovereign identity (Bai, 2022). By additionally adopting a Distributed Ledger Technology (DLT), a temporal immutable history of events and attestations about credentials (DID documents) and actors (DIDs) can be persistently timestamped and proven without a trusted third party or central authority. The decentralized aspect of DLT provides a means of redundant storage of timestamped credential attestations that cannot be mutated by anyone except the owner of the attestation. In addition, a form of Content Addressable Storage (CAS) could be used to also persist the credential information (apart from the attestation metadata). This is to mitigate the risk of the learner or subject from losing their credential document and, to alleviate NLRD system administration and maintenance efforts.

The job markets in both SA and the United States (US) are encountering various challenges and there seems to be a push to create newer types of digital credentials at the risk of credential proliferation (Hurder, 2020). This research assumes that mechanisms are in place to prevent this proliferation and rather focuses on the selection and use of emerging technologies and standards to assist institutions such as SAQA in reducing their operational costs and to facilitate future interoperability. The latest South African data, from 2014, shows that there were approximately 600 foreign qualification verifications per day, and an average of 222,410 local NLRD visits or queries per month (Bolton, 2017), and it is indicated that these queries are related to qualification verifications. In 2016, another qualifications verification entity, the Managed Integrity Evaluation (MIE), also verified 8500 qualifications on average per month (Parliamentary Monitoring Group, 2016). If we assume that these numbers can be added together to ballpark estimate the DLT back-end requirements for verifications per second, we arrive at ((222410 + 8500) * 12 = 365) + 600 = 8192 verifications required per day (which is the assumption our proposed solution is based on). According to the Higher Education Management Information (HEMIS) 2017 data, there are an additional 190, 000 thousand tertiary graduates each year in SA that need to be registered or 190000=365 = 520 per day. All this registration and verification could be streamlined by cryptographic verification with a decentralised, persistent identity.

The difficult task at hand, however, becomes the proper selection of the technologies in question that would provide adequate security, performance, capability, and scalability to achieve such a solution. This is challenging due to the disruptive, emerging, and ever-changing nature of the DLT space. Therefore, this research aims to assess various technologies to answer various questions and gain insight towards selecting a credential and skills tracking system using DLT technologies and accompanying technologies. These technologies can support commercialisation at scale, whilst also complying with the emerging W3C VC and DID standards, the General Data Protection Regulation (GDPR) (as we aim for an internationally recognizable system) and the South African Protection of Personal Information Act (POPIA). These standards then act as a selection filter as there are far too many DLT projects to review that do not consider standardization and future interoperability.

The remainder of the paper is structured as follows: Section 2 briefly outlines the method and reasoning behind the selection. Section 3 provides a summary of each of the candidate DLTs related to their ability or inability to satisfy the assessment criteria. Section 4 provides a summary of analysis after the assessment. Section 5 elicits some important requirements necessary for a minimally viable credential tracking solution, and thus present the proposed system requirements. Finally, Section 6 concludes the study.

2. Method

The first requirement for credential verification is to have a means of identifying stakeholders within the skills and credential tracking system. Thus, the DLT selection and assessment process starts off from the perspective of the W3C DID and VC standards, particularly focussing on the DID method registry (Draft Community Group, 2019). The registry lists all the development entities that have defined or implemented a DID method. In summary, we view the candidate or selected DLTs through the lens of the available DID methods. To identify suitable DLT technology candidates in terms of developer adoption, the registry list was sorted by the DLTs that can host the most amount of different DID method implementations. Therefore, the process of selecting DLTs is based on a non-probability sampling method (which was adopted because there are various DLTs, and it limits the scope to a manageable sample size). The DLTs that support the most DID method implementations are Ethereum, Hedera Hashgraph and Hyperledger Indy. Below is a list of all the selected DLTs and the reasons for their selection:

- Ethereum, because of the aforementioned developer community adoption due to it supporting nine DID methods.
- Hedera Hashgraph, because it supports two DID methods, placing it in second place regarding developer adoption related to standardisation.
- Hyperledger Indy (a.k.a. Sovrin), because it is specifically intended for the standard W3C DID and VCs.
- HoloChain, due to its application- or agent-centric, nonblockchain (non-data-centric) approach, which might scale better than data-centric approaches.
- Inter-Planetary File System (IPFS), because it focuses on storage rather than a blockchain, since credentials contained in DID Documents need to be persisted.
- Factom, because it focuses on generically registering any asset (credentials, property etc.) which aims towards broader prototype context expansion.
- IOTA, because it is fee-less and has low-resource requirements nature (Bhandary et al., 2020) since it focusses on the Internet of Things (IoT) context.

To assess these selected DLT technologies and to ease support of the initial prototyping phase and achieve a minimal viable solution, the following Assessment Criteria (AC) was followed:

- **AC-1**: Does it support a threshold signature scheme where a subset of the credential issuer quorum can cryptographically authorise institutions to issue credentials?
- **AC-2:** Scalability in terms of transactions per second? Even though the scalability requirements for this context are low, the public utility nature of DLTs could cause the tragedy of the commons where others overuse the system, causing it to become slow.
- **AC-3:** How many active addresses there are to indicate developer community and public adoption?
- **AC-4**: Are there documentation examples that illustrate DID document usage to sketch an idea of DID method maturity?
- **AC-5**: Does it feature encrypted DID document storage as well, or does it only store hashes of DID documents on a ledger?
- **AC-6**: Does it have a test or mock-setup for development purposes to avoid paying token costs during testing?

To address these questions, an analysis is performed against the 7 selected DLTs and the results are depicted in section 4.

3. DLT Technologies

This section provides details about the DLT technologies selected in section 2 for the assessment.

3.1. Ethereum

Ethereum is a permissionless, opensource blockchain that features multiple consensus mechanisms and is open to anyone who wishes to participate in its ecosystem. Due to the complexity, measuring the throughput and latency of such a network is difficult as this metric is a function of many parameters and variables along with multiple consensus mechanisms. A best-case scenario experiment from a formal analysis by (Schäffer et al., 2019) which centralises the network to one node, indicates: *"Our experiments show that with a block period of 1s, a block size large enough to fit 1 000 transactions into the block, an Amazon Web Services (AWS) Elastic Compute Cloud (EC2) instance of type c5.4xlarge, and a network of a single node, the throughput can be as high as 328 transactions per second (tps) on average".* From practical measurements, Ethereum can currently process approximately 15tps (Rankhambe & Khanuja, 2019) which might be problematic if decentralized application development gains momentum and, because of the many non-native tokens that are hosted on the Ethereum platform. However, it seems that progress has been made considering Ethereum Improvement Proposal (EIP) #2028, which will provide 9000tps when executing smart contracts and 18,000tps for regular ledger transaction verifications, according to a recent announcement (Dalvit, 2020).

For the period from January 2018 to July 2020, the approximate median of the number of active Ethereum addresses were 500,000 (Bitinfocharts, 2020). Ethereum is also the second biggest market cap platform, which should count in its favour in terms of community. This, in conjunction with the 9 DID method implementations supported by the Ethereum platform, ranks it first for prototype development purposes. Although the *did:signor* DID method is listed as being capable of running on nine different DLTs, no additional documentation or examples were found, thus showing lack of documentation. As of July 2020, there are only two Ethereum associated DID methods that stood out in terms of documentation maturity along with privacy and security considerations, namely, *did:jolo* and *did:selfkey* by jolocom.io (Jolocom, 2019) and selfkey.org (SelfKey, 2017) respectively. These two Self Sovereign Identity (SSI) solutions also aim for GDPR compliance and user-centric data and identity control and provide technical detail encapsulations with open-sourced software libraries. SelfKey is an SSI ecosystem that uses Ethereum as their DID persistence mechanism, but stores credential DID documents on the user's device with more options for DID document storage planned for future development, including integration with a Trezor hardware wallet to secure private keys (SelfKey, 2017). Jolocom is an SSI protocol that also uses Ethereum as its DLT for timestamping attestations and provides DID document persistence via the IPFS CAS as their DID document storage mechanism, which is the same format used to store academic credentials for this solution context. They feature a flexible design with future additions planned to choose different DID document storage back-ends along with IoT device identity management. The white-paper has a better focus on VCs where they aim to satisfy all of Christopher Allen's SSI requirements and also have proper VC and DID documentation examples (Jolocom, 2020). The ten requirements for SSI stipulated by Christopher are: user centricity, control, access, transparency, longevity, portability, interoperability, consent, minimized data disclosure, and protection (Allen, 2020).

3.2. Hedera Hashgraph

Hedera Hashgraph is a permissioned DLT with 12 nodes spread out predominantly across the US and Europe as of July 2020 (DragonGlass, 2020). Depending on the number of nodes, geographical regions and transaction size, Hashgraph's throughput varies from 4,000 to 250,000tps with a latency variation of 20 to 0.04 seconds (Baird & Luykx, 2020). It shows real-world practical use but, in the permissioned setting where nodes are hosted by trusted entities and enrolled through a more formal process. As of 30 July 2020 there were 41,515 active accounts and 217,825,175 transactions processed since 17 August 2019, equating to an average usage of 7.25tps concerning its current real-world throughput demand (DragonGlass, 2020).

The Hashgraph *did:hedera* method is accompanied by many guides (Hashgraph, 2020c)(Hashgraph, 2020a)(Hashgraph, 2020d) and software examples. It is intended to enable the developer to have granular configurability and flexibility regarding their application at the cost of more development or prototyping overhead. The Hashgraph ecosystem features the Hedera Consensus Service (HCS) for token transactions and timestamping of DID associated attestations and includes an integrated Hedera File Service (HFS) to store credentials in DID document format. Application actors can communicate by sending authenticated messages over the HCS in DID document format. All DID documents can be encrypted or unencrypted and allows access control per business application where the servers that host the business application network are registered in an address book which resides in the HFS of the Hashgraph DLT (Hashgraph, 2020b).

DIDs and verifiable credential DID documents can be submitted to a topic identifier and consequently grouped per business application. DID topic access can be controlled by signatures which also support threshold signatures (Hashgraph, 2020b). These signatures achieve a core requirement for the academic AA mentioned in section 5.

3.3. Hyperledger Indy / Sovrin

Hyperledger Indy is a permissioned DLT which is specifically designed for the SSI DID and VC use case, and was pioneered by the Sovrin Foundation (Li et al., 2020). Their DID method is referred to as did:sov. Hyperledger Indy uses the Plenum Redundant Byzantine Fault-Tolerant (RBFT) (Hyperledger Architecture Working Group (WG), 1985) which is based on Practical Byzantine Fault-Tolerant (PBFT) (Castro & Liskov, 2002). It applies redundant instances of the protocol to prevent faulty or malicious nodes from degrading the system's performance. No studies were found that measure and analyse the throughput of Indy's Plenum RBFT, but it is suspected it could be lower than PBFT's 1025tps (Hao et al., 2018) due to its additional redundancy that should theoretically trade-off efficiency. Additionally, no details regarding how many DIDs or active addresses there are on the Hyperledger Indy DLT were found, but there seems to be corporate adoption according to (Gubler, 2019). In terms of documentation, their documentation is scattered across various domains making it difficult to discern which sources of information to trust and how to structure one's learning experience as there is much information redundancy as follows:

- wiki.hyperledger.org/display/indy/Documentation+Index
- wiki.hyperledger.org/display/indy/Hyperledger+Indy
- github.com/hyperledger/indy-node
- github.com/hyperledger/indy-plenum
- github.com/hyperledger/indy-crypto
- github.com/hyperledger/indy-sdk
- github.com/hyperledger/indy-hipe
- readthedocs.org/projects/indy-hipe
- indy.readthedocs.io
- sovrin.org

There has, however, been a proposal to improve their documentation as described in Chapter 17 of their Indy Project Enhancements Documentation (Hardman et al., 2019) to:

1. *"Make better documentation that helps users and contributors to more easily understand, use, and contribute to our code."*
2. *"Help maintainers eliminate duplicated or deprecated content and give everyone a way to efficiently index and search all our documentation across all our repositories."*
3. *"Provide new users a clear path on how to implement the Indy code within their projects, driving adoption of the project and lowering developer burnout."*

There is an extensive walk-through on how to use their higher level VC exchange library named Libvcx (Kulic, 2019). Hyperledger provides tutorials for various programming languages in the Hyperledger Indy Software Development Kit (SDK) documentation that provides granular steps to follow to start developing, but the tutorials were difficult to interpret. A better approach would have been to present a fully working example repository per language with code comments instead of links to individual source files that are wrapped in markdown (Boyd, 2019). Although, this might have been done to reduce documentation efforts on their side. From Chapter 5 of the Indy Project Enhancements Documentation (Hardman et al., 2019), Hyperledger Indy aims to wrap a secret encryption layer around a pluggable storage layer to enable various options for information or DID document storage. However, there seems to be only one storage mechanism, namely RockDB that is used to store credential data (Boyd & Bakov, 2019).

3.4. Inter-Planetary File System (IPFS)

IPFS is a CAS protocol designed to create a permanent, decentralised method of data storage and distribution or sharing, without requiring mutual trust between nodes. IPFS aims to transform the Internet from being a location-based to being a content-based distributed file network and offers the following properties:

- Eliminating the Hypertext Transfer Protocol (HTTP) problem of broken links as an address will always point to the same content added to the IPFS network, because even a slight change in the content will result in a different address.
- Providing censorship resistance considering that web content is not dependent on a single entity.

IPFS has been widely advertised as the new *"permanent web"*, which refers to the permanent reference of the content to which an IPFS address points. Frequently, IPFS is combined with blockchains to store off-chain the actual files while maintaining in the blockchain only the hash-based pointers (or timestamped attestations) to those files (Politou et al., 2020). IPFS synthesises innovative ideas from prior peer-to-peer (P2P) systems, including:

- Distributed Hash Table (DHT) as implemented in the Kademlia protocol for the coordination and maintaining of metadata (Politou et al., 2020).
- BitTorrent inspired communication protocol, BitSwap, to coordinate networks of untrusting peers (swarms) to cooperate in distributing pieces of files to each other (Cohen, 2003).
- Version Control Systems (Git) for supporting file versioning and efficient distribution (Politou et al., 2020).
- Self-certifying File System (SFS) technique for server authentication and to establish a secure communication channel to remote file systems (Politou et al., 2020).

The IPFS DID method (*did:ipid*) supports DIDs on the public and private IPFS networks. It utilises the Interplanetary Linked Data (IPLD) suite, which is a set of tools for describing links (represented in JavaScript Object Notation (JSON)) between content-addressed data, such as IPFS files, Git commits, or Ethereum blocks (libp2p, 2020a). To achieve this, IPLD depends on Content Identifiers (CIDs) for content addressing which is a self-describing, flexible, and an interoperable way of expressing cryptographic hashes. It utilises various multi-formats to accomplish a flexible self-description, namely multi-hash for hashes, multicodec for data content types, and multi-base to represent the base encoding of the CIDs itself (IPFS, 2018). The *did:ipid* DID method also utilises the Inter-Planetary Name System (IPNS) for creating and updating mutable links to IPFS content. The method has minimalistic design goals; a DID trust anchor based on the IPFS and Libp2p protocol (a framework and suite of protocols for building peer-to-peer network applications). A repository exists containing the libp2p specifications that are independent of language or implementation, including wire protocols, addressing conventions, and other "network level" concerns (libp2p, 2020b). The specifications repository serves as a coordination point and a venue to drive future developments in libp2p. Today, implementations of libp2p exist in several languages, with varying degrees of completeness, and the most complete implementations are in Go and JavaScript, with Rust support maturing rapidly. The community is actively working on implementations in python and the Java Virtual Machine (JVM) via Kotlin (libp2p, 2020a). To further enhance security, blockchains and other DLTs could be utilised to anchor the artefacts of the DID method (IPFS, 2018). Currently, asymmetric cryptographic primitives, Rivest–Shamir–Adleman (RSA) and Edwards Elliptic Curve 25519 (Ed25519) are supported, and there are plans to support the elliptic curve used in Bitcoin namely, secp256k1.

3.5. IOTA

IOTA is a DLT that permits hosts in a network to transfer immutable data among each other. It is designed for the IoT industry, which provides secure communications and payments between IoT devices (Foundation, 2020). IOTA's underlying consensus protocol is Tangle, a consensus-building data structure made of a Directed Acyclic Graph (DAG). In the IOTA DAG, graph vertices represent transactions and edges represent approvals. Publishing a transaction in IOTA requires linking a new transaction to any two previous transactions and validating their transaction data. This approach addresses two major issues presented by traditional blockchain-based DLTs, i.e., latency and fees. IOTA offers fast validation, and no fees are required to add a transaction to the tangle. All participants in the network play the same role of issuing and validating transactions and are equally responsible for the consensus (unlike other blockchains where miners are required to validate transactions). Therefore, the cost of a transaction involves only the computational cost of validating two other transactions (Silvano & Marcelino, 2020). IOTA has a throughput of 1500tps with a 1-5minutes or longer transaction time (Foundation, 2020).

IOTA offers Masked Authenticated Messaging (MAM), a communication protocol that includes the functionality to emit and access an encrypted data stream over their Tangle consensus protocol (Zichichi et al., 2020). These streams assume the form of channels, i.e., a linked list of ordered transactions. Once a channel is created, only the owner can publish encrypted messages on it and users in possession of the MAM channel encryption key are authorized to decode the message. MAM also enables users to subscribe and follow a stream of data, generated by some device (Zichichi et al., 2020). The TangleID DID method referred to as did:tangle is intended to implement DIDs and DID documents whilst optimising MAM for key management and related features across the Tangle. The owner of seed in MAM can create a channel structure to transfer the messages. TangleID stores and manages corresponding DID documents on the MAM channels, and uses the initial channel-id as the DID's idstring, whereby each revision of the DID document is recorded on the message of the endpoint afterwards (Su & Wei, 2019). Currently, it can support either Tangle on Mainnet or Tangle on Devnet. There is also a possibility of building on top of Bee (an IOTA Control agenT (ICT)), as long as the interfacing module is complete and a repository is available for further details (Bee, 2020). To create a unique tangle DID, an initial channel needs to be generated with a Merkle-tree signature scheme on top of Winternitz onetime signatures (Su & Wei, 2019).

3.6. Holochain

Holochain is an alternative approach to blockchain and is an open-source framework used to build distributed applications in a P2P network (Holochain, 2020a). Similar to IPFS, it uses a combination of technological techniques (DHT, Gitbased content versioning, digital signatures, peer validation and a gossip protocol) to retrieve and manage its distributed storage. Holochain maintains substantial storage space and network bandwidth, making the system more scalable than a blockchain. This is achieved because Holochain requires each peer to keep its own data within its local storage and each peer is not required to synchronise its own data with all peers in the network (Frahat et al., 2019). Nevertheless, some nodes are responsible for backups to ensure that data is available in case the owner of the data goes offline. Considering the speed of retrieving data, the DHT technique used by Holochain speeds up retrieving data since the data processing is distributed between multiple peers participating in the network (HOLO, 2018). Holochain is not dependent on a global leader consensus, thus limiting the use of computing power (Janjua et al., 2020).

The Holochain *did:holo* method provides examples to assist developers with prototyping and provides details on how to ensure privacy and data security. To run a Create, Read, Update, Delete (CRUD) operation, one must set up local DeepKey instance (*How to Setup DeepKey on Multiple Devices*, 2020) and make Application Programming Interface (API) calls to a Holochain conductor as documented in their developer documentation (Holochain, 2020b). Holochain is a lightweight P2P framework with improved performance characteristics than a *"traditional"* blockchain, i.e., Bitcoin or Ethereum. A write operation to Holochain's DHT takes less than 2 seconds to be accepted whilst key generation takes about 15 seconds (Ulahanna et al., 2019). Unlike most blockchain-based or blockchain-derived projects, Holochain does not have a set tps because it does not have a central point through which all transactions must pass. Instead, Holochain is a generalized protocol for distributed computing with limitless scalability (Forum, 2019).

3.7. Factom

The Factom blockchain is a decentralised publication protocol for building record systems that are immutable and independently verifiable. Factom is built to house data, it exists as a layer above Bitcoin and Ethereum blockchains, and anchors into both every ten minutes. In theory, an attacker would have to compromise Factom, Bitcoin, and Ethereum all at the same time to alter the records, which might be nearly impossible (West, 2020). The cost of using the Factom protocol is a fixed $0.001 per kilobyte (KB) entry and unlike other blockchains, block size is unlimited. Like most other blockchains, Factom does not have the limitation to store everything within a transaction context. According to the Factom real-time explorer, all chains, entries, and transactions are caught up within 1–2 seconds, meaning one need not wait for block confirmation, you can see or share your transactions or data entries instantly after submission to the network. Factom employs a dual-token mechanism which further protects data integrity, and these are:

- Factoids (FCT) - coins that are used to decentralise the network and prevent spam by users. They carry a variable value in relation to the U.S. dollar. They are rewarded to the platform's Authority Node Operators (ANOs) in return for running the protocol's servers and validating new data blocks.
- Entry Credits (EC) - carry a fixed price of one-tenth of a U.S. Penny ($.001) and can be purchased by organisations, in return for storage space on the system. One EC allows an entity to write up to

1KB of data to the blockchain. They have no monetary value and can be purchased with Factoids or with any currency, but only through the Factoid platform (Platform, 2020). Entry Credits are non-transferable and are assigned to one public key on the chain. Therefore, organisations can overcome any stipulations against holding or transacting in cryptocurrencies.

The Factom DID method referred to as *did:factom* describes the low-level data structures and rules for DIDs, DID documents, resolution and registration on Factom itself. Currently, it only supports Factom *"mainnet"* and *"testnet"*, but can be extended to support any number of public or private Factom networks (LLC et al., 2019). This method provides concise examples for DID documents; however, no documentation exists for implementations. The fixed low price data entry means that DIDs also have a fixed low price on Factom. DIDs are primed to become the standard identity solution on top of the Factom protocol, mostly replacing the native identities and replacing the so-called node or server identities that are in place today on the Factom blockchain.

4. Analysis

This section provides a concise analysis based on the above discussed DLT technologies. From understanding the technologies, we can now compare all the assessed technologies and find the most suitable by comparing them against the stipulated criteria. During this qualitative assessment, it was observed that all the DLT technology candidates feature a test or mock setup, usually in the form of a test-net or dev-net to avoid paying unintended testing and prototyping costs. This is assessment criteria (**AC-6**) presented in section 2 and the results are depicted in Table I. The remainder of the criteria are listed in Table I where a ".." indicates a non-definitive "*No*", as no information could be found related to this criterion. Where there was definitive information indicating the lack of a particular criterion, a "*No*" is presented in the table. A "*Yes*" indicates the satisfaction of a criterion associated with a particular DID method and DLT.

Table 1. DLT Assessment Criteria Measurement Matrix

DLT/DID Method	**AC-1**	**AC-2**	**AC-3**	**AC-4**	**AC-5**	**AC-6**
did:jolo	..	..	Yes	Yes	Yes	Yes
did:selfkey:	..	..	Yes	Yes	..	Yes
did:hedera	Yes	Yes	..	Yes	Yes	Yes
did:sov	..	..	Yes (corporate)	Yes	..	Yes
did:ipid	..	No (Huang et al., 2020)	..	..	..	Yes
did:tangle	..	..	..	..	Yes	Yes
did:holo	..	Yes	..	..	Yes	Yes
did:factom	..	..	..	..	..	Yes

Out of the seven selected DLT technologies, Table I shows that, when compared against the assessment criteria presented in section 2 (consisting of six criterion), only Hedera Hashgraph meets five of the requirements. The other technologies, mostly fall under the undefined category "..". However, because we aim to build the system now, the "*undefined*" is treated as a *"no"* for this assessment. It is worth noting that, because of the improvements towards DLTs, some of the criterion might be met soon. Thus, the most scalable candidate with proper quantitative practical scalability measurements is Hedera Hashgraph, although, it is more centralised due to its permissioned network configuration.

The second candidate in terms of assessment criteria satisfaction is Jolocom which uses IPFS for DID document persistence and Ethereum as DLT for attestation anchoring or timestamping. Although it doesn't have threshold signature capabilities yet, it might support it by the time of the development phase, and it will be monitored going forward because it has good developer documentation and examples. All the other DLTs and DID method candidates seem to be lacking, which indicates the new and emergent nature of the DLT space and the inability for them to satisfy the requirements to be outlined in Section 5 in their current state, particularly for rapid prototyping purposes.

5. The proposed solution requirements

In this section, we define a conceptual proposed solution requirements that abstracts a decentralised academic and skills credentialing system. For this proposed system, we plan to incorporate a DLT, W3C VC and DID standards. This paper presented a DLT based assessment, and with that, we can continue with the development of the system. However, some requirements for using the W3C standards needs to be addressed, which is accomplished by this section. More details about the standards, but not the requirements for the intended system, are provided in this paper (Pretorius et al., 2021). There are three main players for the W3C VC data model (issuer, holder and verifier) which are depicted in Figure 1.

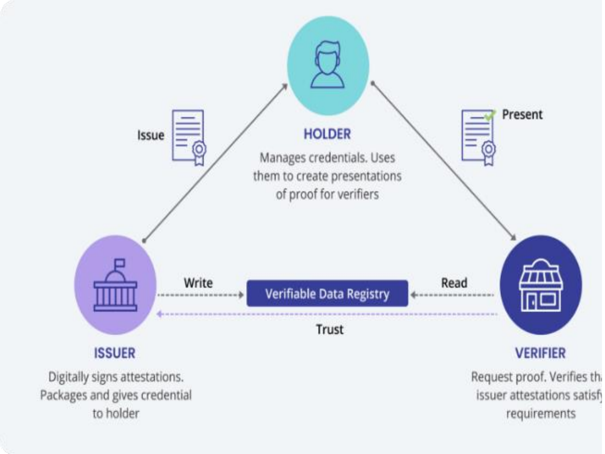
HOLDER
Manages credentials. Uses them to create presentations of proof for verifiers
Issue
Present
Write
Verifiable Data Registry
Read
Trust
ISSUER
Digitally signs attestations. Packages and gives credential to holder
VERIFIER
Request proof. Verifies th
issuer attestations satisf
requirements

Figure 1. W3C VC data model ecosystem overview (Affinidi, 2021)

To consider what is needed for a minimally viable credential tracking and verification solution, we conducted research and collected several requirements which are discussed in the following sections. From (Gräther et al., 2018) and (Gresch et al., 2019), it was established through stakeholder engagements and interviews with certification authorities that the requirements can be classified into three predominant sections, namely: issuer-, verifier- and holder or subject requirements. We refer to these requirements in shorthand as Issuer Requirements (IR), Verifier Requirements (VR) and Holder Requirements (HR).

5.1. Issuers

Issuer requirements relate to the AA quorum of size k and the certificate issuing entities such as universities or online educators. The AA must issue an accreditation credential to each university or educator, and each educator then issues the academic or skill-set achievement credentials towards the learners. Only a to-be-determined *n* of *k* signatures will be required to issue the accreditation credential towards institutions. Below are the requirements:

- **IR-1**: The system should allow only accredited certificate issuers to generate certificate credentials (Gräther et al., 2018) and (Gresch et al., 2019), and enable them to search, browse as well as list learner DIDs or issued credentials and examination results associated with a learning course (Gräther et al., 2018).
- **IR-2**: The solution should import credential data and examination results from, for example, the SAQA NLRD (Gräther et al., 2018). The credentials must be digitally signed using the issuer private key and registered on a DLT during the import phase. The credential must be stored in a decentralized CAS (Gräther et al., 2018).
- **IR-3**: The solution should allow certification authorities to queue, sign, issue and timestamp credential information in bulk, or one-by-one on a DLT (Gräther et al., 2018).
- **IR-4**: The solution should allow certification authorities to revoke the credential certificate when plagiarism was detected, or when the credential expiry date has been reached (Gräther et al., 2018).

5.2. Verifiers

Verifier requirements relate to entities such as employers, job recruiters, SAQA and anyone who wishes to verify a credential related to a DID since all actors will require at least one DID to participate in this system and these are the requirements:

- **VR-1**: The solution must allow any system actor (issuer, verifier, or subject) to verify the authenticity of any credential by looking up the timestamped metadata on the DLT, as well as verifying the cryptographic issuer and holder digital signatures. The solution should provide the ability to select and queue multiple credentials to be verified in bulk (Gräther et al., 2018).
- **VR-2**: The verification process and interface should be as automated as possible and hide technical details (Gresch et al., 2019).

5.3. Holders or Subjects

Holder requirements relate to both the distributed CAS and learner (a.k.a. the *subject*) and puts the learner in control of their data which is in alignment with the GDPR requirements:

- **HR-1**: The solution must allow credentials to contain an event list that will notify a list of associated DIDs or actors when certain credential access events occur, e.g., credential expiry, credential read- and verification events. This notification list should be configurable by the learner or subject, which is also the holder. The credential should also contain default notification events to notify the issuer and learner when the credential expires, unless it is a permanent certificate (Gräther et al., 2018).
- **HR-2**: The solution must allow learners to create, manage and share job application portfolios with other DID identified actors. This requirement seems related to W3C verifiable presentations, which should prevent anyone from copying information from this view as it is read-only information (Gräther et al., 2018).
- **HR-3**: The solution must notify learners when verifiers read or verify their credentials after a credential or verifiable presentation has been shared (Gräther et al., 2018). This could be accomplished by maintaining credential stateful metadata within a distributed and encrypted CAS mechanism in the form of another DID document.

With these requirements, we can then ensure that the prototype is in line or meets all the stipulated requirements for it to be accepted as a viable solution. From the analysis, it was discovered that from the eight selected DLTs, Hedera Hashgraph was the most appropriate for the intended system when compared against the requirements. Therefore, it is worth noting that Hedera Hashgraph also can satisfy prime requirement **IR-1** (built-in threshold signature capability) above, and has proper documentation and software development guides. However, it is worth noting that some of the requirements can only be tested during the implementation process. The *did:hedera* DID method and associated back-end is therefore our primary candidate to realise this skills tracking and credentialing solution and ensure that the above mentioned requirements are met.

6. Conclusion and future work

In this paper, we focus on how DLT technologies can be a key technology to enable academics and skills credentialing, tracking and verification system. Various DLT technologies have been introduced over the years since inception. With that, it is vital to choose a DLT technology that meets the requirements of the proposed system. Therefore, this paper provides an assessment of DLT technologies that have been picked in a non-probability sampling method. To successfully conduct the assessment, the seven selected DLTs were assessed against the assessment criterion discussed in section 2. Furthermore, the proposed solution is presented which aims to utilise VCs, DIDs and W3C; and, with that, certain requirements have been discussed in section 5. These requirements go into detail as to what is required for a viable credentialing or skills tracking system. With the proposed solution already in place, and results from the assessment, the remaining thing is to employ the selected DLT technology to develop a demonstratable prototype. DLTs are gaining momentum and new improvements are introduced frequently. A baseline assessment for a skills tracking system was presented, and there is a possibility for change in future. Thus, it is vital to remain updated with these improvements.

7. References

Affinidi. (2021). *What are Verifiable Credentials (VCs), Demystified.* https://academy.affinidi.com/what-are-verifiable-credentials-79f1846a7b9

Allen, C. (2020). *The Path to Self-Sovereign Identity*. https://www.good-id.org/en/articles/path-self-sovereign-identity/

Bai, Y. (2022). *Decentralized and Self-Sovereign Identity in the Era of Blockchain : A Survey. 2021*, 500–507. https://doi.org/10.1109/Blockchain55522.2022.00077

Baird, L., & Luykx, A. (2020, August). The Hashgraph Protocol: Efficient Asynchronous {BFT} for High-Throughput Distributed Ledgers. *2020 International Conference on Omni-Layer Intelligent Systems ({COINS})*. https://doi.org/10.1109/coins49042.2020.91

91430

Bee. (2020). *BEE: A framework for building IOTA nodes, clients, and applications in Rust.* GitHub. https://github.com/iotaledger/bee

Bhandary, M., Parmar, M., & Ambawade, D. (2020, June). A Blockchain Solution based on Directed Acyclic Graph for {IoT} Data Security using {IoTA} Tangle. *2020 5th International Conference on Communication and Electronics Systems ({ICCES}).* https://doi.org/10.1109/icces48766.2020.9137858

Bitinfocharts. (2020). *Ethereum Active Addresses chart*. https://bitinfocharts.com/comparison/ethereum-activeaddresses.html

Bolton, H. (2017). *2014 Assessment of the Impact of the South African National Qualifications Framework: FULL REPORT*. South African Qualifications Authority (SAQA). https://saqa.org.za/sites/default/files/2019-11/2017 04 20 IS v3_Part1_0.pdf

Boyd, M. (2019). *Hyperledger Indy How To Tutorials*. https://github.com/hyperledger/indy-sdk/tree/master/docs/how-tos

Boyd, M., & Bakov, A. (2019). *Hyperledger Indy Storage Components*. https://github.com/hyperledger/indy-plenum/blob/master/docs/source/storage.md

Castro, M., & Liskov, B. (2002). Practical byzantine fault tolerance and proactive recovery. *{ACM} Transactions on Computer Systems*, *20*(4), 398–461. https://doi.org/10.1145/571637.571640

Cohen, B. (2003). Incentives build robustness in BitTorrent. *Workshop on Economics of Peer-to-Peer Systems*, *6*, 68–72.

Dalvit, L. (2020). *Ethereum: a new record of transactions per second*. https://en.cryptonomist.ch/2020/01/07/ethereum-transactions-per-second-istanbul/

Draft Community Group. (2019). *DID Method Registry: A registry for Decentralized Identifier Methods*. https://w3c-ccg.github.io/did-method-registry/

DragonGlass. (2020). *Live and Historical data for Hedera Hashgraph*. https://app.dragonglass.me/hedera/home

Forum, H. (2019). *What is the TPS (Transactions Per Second) on Holochain?* https://forum.holochain.org/t/what-is-the-tps-transactions-per-second-on-holochain/191

Foundation, A. (2020). *What Is The Fastest Blockchain And Why? Analysis of 43 Blockchains*. https://alephzero.org/blog/what-is-the-fastest-blockchain-and-why-analysis-of-43-blockchains/

Foundation, I. (2020). *IOTA*. https://www.iota.org/

Frahat, R. T., Monowar, M. M., & Buhari, S. M. (2019). Secure and Scalable Trust Management Model for {IoT} P2P Network. *2019 2nd International Conference on Computer Applications {\&} Information Security ({ICCAIS})*, 1–6. https://doi.org/10.1109/cais.2019.8769467

Gräther, W., Kolvenbach, S., Ruland, R., Schütte, J., Torres, C., & Wendland, F. (2018). Blockchain for Education: Lifelong Learning Passport. *European Society for Socially Embedded Technologies (EUSSET)*, *2*(10). https://doi.org/10.18420/BLOCKCHAIN2018_07

Gresch, J., Rodrigues, B., Scheid, E., Kanhere, S. S., & Stiller, B. (2019). The Proposal of a Blockchain-Based Architecture for Transparent Certificate Handling. In *Business Information Systems Workshops* (pp. 185–196). Springer International Publishing. https://doi.org/10.1007/978-3-030-04849-5_16

Gubler, S. (2019). *Hyperledger Indy Graduates To Active Status; Joins Fabric And Sawtooth As "Production Ready" Hyperledger Projects*. https://www.hyperledger.org/blog/2019/04/10/hyperledger-indy-graduates-to-active-status-joins-fabric-and-sawtooth-as-production-ready-hyperledger-projects

Hao, Y., Li, Y., Dong, X., Fang, L., & Chen, P. (2018, June). Performance Analysis of Consensus Algorithm in Private Blockchain. *2018 {IEEE} Intelligent Vehicles Symposium ({IV}).* https://doi.org/10.1109/ivs.2018.8500557

Hardman, D., Curren, S., & George, N. (2019). *Indy Project Enhancements Documentation, Hyperledger Indy*. https://readthedocs.org/projects/indy-hipe/downloads/pdf/latest/

Hashgraph. (2020a). *Decentralized Identifiers: User Guide*.

https://github.com/hashgraph/did-sdk-java/blob/master/docs/did-user-guide.md

Hashgraph. (2020b). *Hedera Hashgraph DID Method Specification: Version 0.1*. https://github.com/hashgraph/did-method/blob/master/did-method-specification.md

Hashgraph. (2020c). *Identity Network: User Guide*. https://github.com/hashgraph/did-sdk-java/blob/master/docs/id-network-user-guide.md

Hashgraph. (2020d). *Verifiable Credentials Registry: User Guide*. https://github.com/hashgraph/did-sdk-java/blob/master/docs/vc-user-guide.md

HOLO. (2018). *Here's Holochain in 100, 200, and 500 words*. Medium. https://medium.com/h-o-l-o/heres-holochain-in-100-200-and-500-words-509818aa3c88

Holochain. (2020a). *Holochain*. https://holochain.org/

Holochain. (2020b). *Holochain Guidebook*. https://developer.holochain.org/docs/guide/welcome/

How to Setup DeepKey on multiple devices. (2020). HackMD. https://hackmd.io/xp5h1bkLRy-oef45tiJ1yw

Huang, H., Lin, J., Zheng, B., Zheng, Z., & Bian, J. (2020). When Blockchain Meets Distributed File Systems: An Overview, Challenges, and Open Issues. *{IEEE} Access, 8*, 50574–50586. https://doi.org/10.1109/access.2020.2979881

Hurder, S. (2020). *Blockchain Tech Can Verify Credentials, but Beware Credentialism*. https://web.archive.org/web/20201101010813/https://www.coindesk.com/blockchain-based-verifiable-credential-proliferation

Hyperledger Architecture Working Group (WG). (1985). *Hyperledger Architecture, Introduction to Hyperledger Business Blockchain Design Philosophy and Consensus* (Paper, Vol. 1). Hyperledger.org. https://www.hyperledger.org/wp-content/uploads/2017/08/Hyperledger_Arch_WG_Paper_1_Consensus.pdf

IPFS. (2018). *IPID DID Method*. GitHub. https://did-ipid.github.io/ipid-did-method/

Janjua, K., Shah, M. A., Almogren, A., Khattak, H. A., Maple, C., & Din, I. U. (2020). Proactive Forensics in {IoT}: Privacy-Aware Log-Preservation Architecture in Fog-Enabled-Cloud Using Holochain and Containerization Technologies. *Electronics, 9*(7), 1172. https://doi.org/10.3390/electronics9071172

Jolocom. (2019). *A Decentralized, Open Source Solution for Digital Identity and Access Management*. https://jolocom.io/wp-content/uploads/2019/12/Jolocom-Whitepaper-v2.1-A-Decentralized-Open-Source-Solution-for-Digital-Identity-and-Access-Management.pdf

Jolocom. (2020). *Credentials \& Signed Credentials — Jolocom-Lib documentation*. https://jolocom-lib.readthedocs.io/en/latest/signedCredentials.html

Kulic, D. (2019). *Getting Started with Libvcx: A Developer Guide for Building Indy Clients Using Libvcx*. https://github.com/hyperledger/indy-sdk/blob/master/vcx/docs/getting-started/getting-started.md#what-indy-libindy-and-libvcx-are-and-why-they-matter

Li, D., Wong, W. E., & Guo, J. (2020, January). A Survey on Blockchain for Enterprise Using Hyperledger Fabric and Composer. *2019 6th International Conference on Dependable Systems and Their Applications ({DSA})*. https://doi.org/10.1109/dsa.2019.00017

libp2p. (2020a). *libp2p documentation portal*. https://docs.libp2p.io/

libp2p. (2020b). *libp2p specification*. GitHub. https://github.com/libp2p/specs

LLC, F., BV, S., & Inc, F. (2019). *Factom Decentralized Identifiers (DID)*. https://github.com/factom-protocol/FIS/blob/master/FIS/DID.md

Ntshangase, B. A., & Msosa, S. K. (2022). *Research in Business & Social Science The role of the South African qualifications authority in curbing misrepresentation of qualifications*. *11*(6), 421–429.

Parliamentary Monitoring Group. (2016). *Annexure A'': National Qualifications Framework (NQF) Amendment Bill*. https://static.pmg.org.za/190220Annexure_A_to_MIE.pdf

Platform, H. (2020). *Factom (FCT) Gets Listed On Halodex Exchange!* https://medium.com/@haloplatform/factom-fct-gets-listed-on-halodex-exchange-f20ab0896bc6

Politou, E., Alepis, E., Patsakis, C., Casino, F., & Alazab, M. (2020). Delegated content erasure in {IPFS}. *Future Generation Computer Systems*, *112*, 956–964. https://doi.org/10.1016/j.future.2020.06.037

Pretorius, M., Dlamini, N., & Mthethwa, S. (2021). Towards Academic and Skills Credentialing Standards and Distributed Ledger Technologies. *ICISSP*, 249–257.

Rankhambe, B. P., & Khanuja, H. K. (2019, September). A Comparative Analysis of Blockchain Platforms {\textendash} Bitcoin and Ethereum. *2019 5th International Conference On Computing, Communication, Control And Automation ({ICCUBEA})*. https://doi.org/10.1109/iccubea47591.2019.9129332

Schäffer, M., di Angelo, M., & Salzer, G. (2019). Performance and Scalability of Private Ethereum Blockchains. In *Business Process Management: Blockchain and Central and Eastern Europe Forum* (pp. 103–118). Springer International Publishing. https://doi.org/10.1007/978-3-030-30429-4_8

Schneier, B. (2019). *Essays: Testimony before the Senate Judiciary Committee*. https://www.schneier.com/essays/archives/2007/05/testimony_real_id.html

SelfKey. (2017). *The SelfKey Foundation*. https://selfkey.org/wp-content/uploads/2019/03/selfkey-whitepaper-en.pdf

Silvano, W. F., & Marcelino, R. (2020). Iota Tangle: A cryptocurrency to communicate Internet-of-Things data. *Future Generation Computer Systems*, *112*, 307–319. https://doi.org/10.1016/j.future.2020.05.047

Sonnino, A., Al-Bassam, M., Bano, S., & Danezis, G. (2018). Coconut: Threshold Issuance Selective Disclosure Credentials with Applications to Distributed Ledgers. *Computing Research Repository (CoRR)*, *abs/1802.07344*. http://arxiv.org/abs/1802.07344

South African Qualifications Authority (SAQA). (2019). *Strategic Plan for 2020/21 – 2024/25*. https://www.saqa.org.za/sites/default/files/2020-04/Strategic Plan 2020_25.pdf

Su, T., & Wei, W. Y. (2019). *TangleID DID Method Specification*. https://github.com/TangleID/TangleID/blob/develop/did-method-spec.md

Ulahanna, J., Brock, A., Sporny, M., Duffy, K. H., Sabadello, M., Zagidulin, D., & Burnett, D. (2019). *did:holo method*. https://github.com/WebOfTrustInfo/rwot9-prague/blob/master/draft-documents/did:hc-method.md

West, A. (2020). *Factom Protocol Offers Cost-Predictive and Compliant Blockchain-Based Services to Enterprises and Large Organizations*. https://web.archive.org/web/20200815125659/https://www.cardrates.com/news/factom-protocol-is-a-robust-blockchain-platform-for-enterprises/

Zichichi, M., Ferretti, S., & D\textquotesingleAngelo, G. (2020, July). On the Efficiency of Decentralized File Storage for Personal Information Management Systems. *2020 {IEEE} Symposium on Computers and Communications ({ISCC})*. https://doi.org/10.1109/iscc50000.2020.9219623