*Annals of Emerging Technologies in Computing (AETiC)*
*Vol. 4, No. 5, 2020*

*Review Article*

# Decentralized Identity Systems: Architecture, Challenges, Solutions and Future Directions

**Omar Dib[1,*] and Khalifa Toumi[2]**
[1]Department of Computer Science, Wenzhou-Kean University, Wenzhou, China
odib@kean.edu
[2]IRT SystemX, Paris-Saclay, France
khalifa.toumi@irt-systemx.fr
*Correspondence: odib@kean.edu

Received: 18th September 2020; Accepted: 10th December 2020; Published: 20th December 2020

**Abstract: Due to the exponential rise of the Internet in the last two decades, the digital presence has seen an enormous increase. Today, billions of people, devices and objects are digitally connected making the interactions much easier than before. To securely establish this connectivity in the digital era, proving digital identities has become crucial. Due to this, a growing number of organizations are building solutions that establish, verify and manage digital identities. Yet, a solution whereby digital identities and their associated data are efficiently managed is still far from being achieved. To fully understand the reason behind this lack, this paper provides a detailed state of the art related to identity management systems. It overviews traditional systems, analyses their strengths and limitations. This work highly focuses on the novel decentralized identity systems based on blockchain; a complete study describing their architecture, components, lifecycle and workflow is detailed. Additionally, solutions enabling decentralized identity are discussed, analysed and compared according to the ten principles of self-sovereign identity. Lastly, the challenges hindering the shift toward the fully decentralized identity paradigm are discussed.**

**Keywords:** ***Blockchain; Digital Identity; Identity Management Systems; Self-Sovereign Identity***

## 1. Introduction

The rise of the Internet and Web 2.0 in the last two decades has led to an evolution in the way interactions are performed between digital entities [1]. The society has become almost fully connected from online banking, e-commerce to messaging and booking travels, practically all sectors are moving their services towards the digital space. This continuous trend towards online applications has dramatically changed the behaviors of businesses and customers. For businesses, a digital oriented economy requires finding new ways to interact with customers; the way businesses market their products have to be totally changed. Businesses’ relationships have not only changed between customers and products, but also with partners, suppliers, and employees.

Relationships are nowadays increasingly moving to the electronic world and being mediated by automated processes rather than intermediary people [2]. For customers, the changes are equally dramatic. Where people used to purchase things from others at physical locations, transactions are currently made using online services. Moreover, the classical trust vectors that most customers have relied on, are either absent nowadays or can be forged.

To account for all those changes, having an efficient management of Digital Identity (DI), and a better control of interactions have become a must. To be defined, a DI is a set of information that is used to represent an entity in the digital world [3]. This entity may be a person, organization,

Omar Dib and Khalifa Toumi, "Decentralized Identity Systems: Architecture, Challenges, Solutions and Future Directions”, *Annals of Emerging Technologies in Computing (AETiC)*, Print ISSN: 2516-0281, Online ISSN: 2516-029X, pp. 19-40, Vol. 4, No. 5, 20th December 2020, Published by International Association of Educators and Researchers (IAER), DOI: 10.33166/AETiC.2020.05.002, Available: http://aetic.theiaer.org/archive/v4/v4n5/p2.html.

application, or even a device. The information comprised in a DI allows for assessing and authenticating an entity on the web, without the involvement of human operators. Such information is stored in computer systems, and usually encompasses the civil or national attributes such as name, date of birth, address, etc.

The DI is currently so widespread that can be referred to by many discussions as the entire collection of data generated by an entity's online activity. This information can be divided into attributes, preferences, and traits [4]. The attributes refer to an entity's acquired data such as the past purchase behaviors of a customer. The preferences represent the entity's desires such as the preferred seating on an airline. Finally, traits are like attributes, features of the entity, but they are inherent rather than acquired such as the birth date.

A DI does not only encompass data that uniquely describes an entity but also information about its interactions and relationships with its environment. An entity may have multiple sub-identities, each containing a subset of attributes that are relevant to a specific topic or used in certain circumstances.

Due to recent advancements in the digital era, many use cases have been enabled where the DI plays a key role. These use cases highly drive the innovation and are essential to build a robust digital world. For instance, the secure logging service has allowed for proving that "someone is the real person he claims he is". This has consequently enabled the access to a wide range of online applications such as bank accounts, customer portals, medical applications, and so on.

Due to this involvement of DI in almost all the online-based applications, individuals and companies are constantly seeking efficient Identity Management Systems (IMSs) across multiple services, markets, standards and technologies. This is crucial for entities while establishing online communications, and for businesses to identify employees, systems, resources, and services.

Despite the progress achieved in the DI domain, more data breaches are still faced; vulnerabilities of weak identity systems are almost daily highlighted. Recent studies showed that nearly 60 millions of Americans were affected by identity theft in 2017 [5]; more than 10 billion records have been breached since 2013 [6]; over 6,500 incidents resulted in compromised data were publicly disclosed in 2018; the average cost of an identity fraud is estimated to $263 per person; the yearly total cost of identity theft was estimated in 2016 to 16$ billion [7]; the average cost for each stolen or lost record containing sensitive and confidential data is estimated to 148$[1]. All those facts undoubtedly require developing novel identity solutions that strive towards a return to a satisfactory level of privacy.

Within the context of DI, this paper aims at providing a constructive state of the art related to existing IMS(s). Their advantageous and limitations are discussed from multiple points of view, including users, businesses, technical components and regulations. This work also discusses the recent advances in the DI field resulting from the arise of novel technology, namely the blockchain. The impact of this new technology is studied in order to understand how a more efficient IMS can be built.

The main question this study focuses on is whether novel IMSs such as the decentralized ones based on blockchain will totally solve traditional identity related issues such as data' security, users' privacy and protection of personal attributes? In order to answer this question, analyzing the limitations of classical IMSs is firstly done. After that, the novel IMSs are presented, and their strengths and limitations are discussed.

The remainder of this article is organized as follows. Section 1 and 2 introduce the topic and context of DI and explain the most used term in this era. In Section 3 centralized identity solutions are discussed. Section 4 presents the novel Self Sovereign Identity (SSI) paradigm that is based on blockchain. Later, some decentralized identity solutions are presented, discussed and evaluated according to the ten principles of SSI. Section 5 details the challenges and limitations of decentralized identity. Section 6 overviews existing surveys and discuss their main differences and lacks compared to this paper. Finally, some conclusions are provided and an outlook of future works is given.

[1] https://selfkey.org/ decentralized-identifiers-article/

## 2. Preliminaries

To get a better understanding of the terms used in this paper, some definitions and abbreviations are explained here in more details.

- **Identity Management System (IMS):** refers to a system or a set of technologies that can be used for enterprise or cross-network identity management.
- **Digital Identity (DI):** may be defined as a set of information that is used to represent, assess and authenticate an entity in the digital world without the involvement of human operators.
- **A Subject or Entity:** is a person, organization, software program, machine or device that is digitally connected. An entity has to hold a DI so that it can be authenticated and eventually authorized in case of requesting access to online resources.
- **Attribute:** is a characteristic of an entity. Attributes can be persistent such as the date of the birth of an individual, temporary (e.g. an address) or long-lived (e.g. social security number).
- **Identifiers:** is defined as a set of difficult or impossible to alter attributes associated with an entity. These attributes are usually referred to as persistent identifiers, and used to identify the entity. An example of such identifiers is the genetic pattern of an individual.
- **Identification:** is the association of identifiers with an entity presenting attributes. Examples include associating a physical person to a claimed name; connecting a company to a financial record; or linking a patient with physical attributes.
- **Authentication:** is the process of providing the appropriate credentials to prove the identity of an entity. For example, when the appropriate user name and password are provided, the user proves the account's ownership. There are several methods of authentication [8], which are:
    1. **Something You Know**, such as a PIN code or a password.
    2. **Something You Have**, such as an identity card, bank card, smart card, security token, mobile phone or an ID document.
    3. **Something You Are**, such as fingerprints, face, irises and voice.
    4. **Something You Do**, such as motor skills, gestures and keystrokes or applications.
- **Authorization:** Once the process of identification and authentication is successfully done, an entity can be granted authorization based on its proven identity. Particular actions can therefore be allowed based on the entity attributes. Examples include a person ability to claim lines of credit or an emergency vehicle right to pass through a red light.
- **SSI: Self-Sovereign Identity:** refers to a new IMS whereby the user should fully own his/her identity data without any intervention from an outside administration.
- **Decentralized IDentifier (DID):** refers to a new type of digital identifiers that enables a verifiable, decentralized DI.
- **Verifiable Credentials (VCs):** refers to the electronic equivalent of the physical credentials such as a passport or a driving license.
- **Proof of Authority (PoA):** is a consensus algorithm that delivers comparatively fast transactions through a consensus mechanism based on identity as a stake.
- **Zero Knowledge Proof (ZKP):** is a digital method whereby one party proves to another party the possession of an information without revealing it.
- **Software Development Kit (SDK):** refers to a collection of software used for developing applications for a specific device or operating system.

## 3. Classical Identity Management Systems

An IMS defines the whole lifecycle of a DI, from creation, storage, authentication, authorization till revocation and destruction. Identity solutions can be classified under three categories: centralized [9], federated [10] and decentralized [11].

The centralized and federated ones refer to classical systems since identity attributes in them are managed by a third party such as an identity provider.

The decentralized system is however the new form of managing digital identities where attributes are managed by the users themselves. In this section, the workflow, benefits and limitations of classical systems are discussed.

### 3.1. Centralized Identity

This is the traditional paradigm in which an individual accesses the services of an organization that manages or own the identity system. The owner of the system collects, stores and uses the individual's identity and its related data. Such systems are currently proposed by private organization such as banks, social media companies, and even governments.

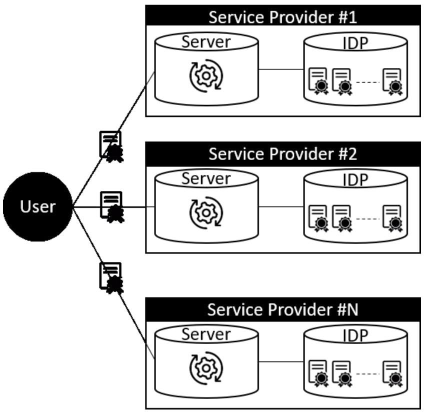

**Figure 1.** Centralized identity model

The adoption of centralized identity systems is nowadays so widespread. Most of authentications are currently established through a matching verification between a login and password. A digital account is usually created by the user and stored within the databases of a service provider. A user usually has one account for each service provider as shown in Figure 1.

The quality of data in such identity systems varies according to the policies of the service provider holding the user's account. For instance, heavily regulated sectors such as governments and financial services have thorough identity proofing process. However, in other systems such as in some social media platforms creating multiple or even malicious identities can be easily done.

It is clear that this approach has successfully enabled a digital representation of an entity, and thus allowed for a wide range of online services [12]. However, since data are managed by a third party, the privacy of users may be compromised and their online activities may be linked and eventually traced. In addition, as the user will have to create one identity per service provider, an extremely fragmented landscape will be created, and non-fluid experience will accompany the user through the online presence. From a service provider point of view, such approach usually requires investing high resources to store, maintain, and protect users' data [13].

### 3.2. Federated Identity

A federated identity system is the result of establishing mutual trust between two or multiple centralized systems [14]. That is done either by distributing the components of verification and trust across all the identity systems, or by mutually accepting the standards applied by each system[2].

For example, international organizations or even governments may agree to accept each other's credentials. That is usually associated with creating common standards such as eIDAS [15] that provides standards for cross-border travels in the European Union. Enterprises may also agree to accept each other's identity proofing systems. The owners of the identity systems usually establish a one-to-one trust via legal agreements and common technical standards. As the number of trusted relationships increases, the network grows and so does its reputation.

[2] http://www3.weforum.org/docs/WEF INSIGHT REPORT DigitalIdentity.pdf

Users often like the convenience that the federated identity provides while accessing multiple services in different platforms. That has led to a wide adoption of federated systems. However, building trust relationships between two or multiple system owners is not always an easy task limiting the fast implementation of that archetype. As with centralized systems, the trust level may highly vary depending on the system owners, the identity verification degree and the data vetting process that they perform.

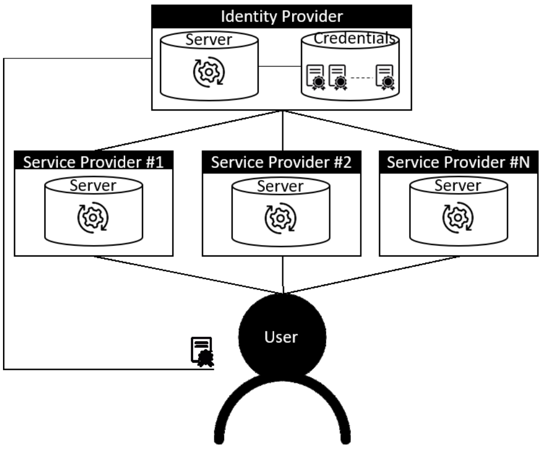

**Figure 2.** Federated identity model

The high-level architecture of the federated identity model is shown in Figure 2. One of the most common federated identity systems used is the third-party authentication API provided by OAuth [16]. Many web services propose to use Google or Facebook accounts to use their services. They rely on the trust that the user's identity has been verified by other party. This authentication method is widely adopted since most online users have a Facebook or a Google account. However, with that method, identity providers were able to trace the users' activities and cross that information with identity data giving them a lot of confidential information[3].

As in centralized systems, individuals in federated platforms might be given little choice over when, where, how and by whom their data are processed. The complexity for systems owners arises from the eventual need for legal agreements, including the liabilities, division of risks, and creation of technical standards. Those complexities may engender high implementation cost which often lead to an absence in many services that users would like to have.

## 4. Towards Decentralized Identity

### 4.1. Definition

[redacted]

**Figure 3.** Decentralized identity model

In the above discussed centralised and federated identity paradigms, the identity is provided by a third party. Oppositely, in the decentralised identity paradigm, the entity itself is placed at the core of exchanges and the need for third parties to manage identities is removed.

[3] . https://www.wired. com/story/security-risks-of-logging-in-with-facebook/

This shift is enabled by putting as much identity infrastructure and information as possible in the entity's hands, and relying on trustworthy decentralised tools and techniques such as cryptographic algorithms and secure distributed ledger to produce and store mathematical proofs about the veracity of identity attributes and their associated data [17]. The general architecture of a decentralized architecture is presented in Figure 3.

A decentralized IMS does not depend on one or even a set of service providers to establish and manage identities. It relies instead on mainly, a private data store to keep the identity attributes, and a digital device to manage their life-cycle. Both components should be managed by the entity itself, otherwise middleware are reproduced. The data store that often takes the form of a mobile phone, personal computer or private cloud storage, also holds attestations from traditional trust anchors, such as governments, banks, employers, etc. The entity chooses via the digital device manager which attestation or data attributes to share and with whom to share it, and for which purpose the share is taking place. By doing so, the entity is supposed to keep control over its identity attributes and associated attestations.

A decentralized identity is supposed to be safe as long as the entity in possession of the identity keeps control over it. This means that the entity assumes responsibility for its identity data. For many users, this trade-off seems to be acceptable.

In a decentralized identity world, the entity creates its own digital identity(ies). This usually begins with the creation of one or several unique identifiers, and then attaching authentic and provable attributes to them. Once this is done, the entity can collect credentials from trusted anchors and make them available when needed. A typical use case would be for a person to collect credentials from his/her university. When the claim is needed, for example to apply for a student loan, the person can simply present the appropriate credential.

To prove the authenticity of a credential (i.e. that is issued by the named authority and has not been tampered with since, and that the person who presents it is the same person being referred to), various cryptographic techniques such as digital signatures can be applied.

Currently, the term Verifiable Claims or Verifiable Credentials (VCs) is used to refer to such digital credentials that can be cryptographically verified. A VC does not only give a user much more control over his or her identity attributes, it also makes the DI much easier to use. More importantly, once issued, a VC can be easily transferred with many service providers and employed on multiple websites.

Additionally, when a VC changes, for instance if the user's address has changed, the modification needs to only be registered once; that is, the user locally makes the change and then automatically broadcasts it to his/her connected service providers.

A decentralized identity is not only supposed to improve the users' digital experience, one of its promises is also to facilitate the process for businesses, which have been trying for many years now, but without success, to get rid of the complex, costly and risky identity management process.

A decentralized identity will not completely get rid of intermediate parties. It will, however, still, and to a large extent be reliant on third parties to provide signed data that will be transformed to portable VCs. Like their physical counterparts, a VC will remain under the control of the authority that issued it (e.g. the state can issue a driving license, and can also revoke it). For many use cases, relying on trusted third parties to issue VCs that can be associated with a user-generated identifier would be desirable.

## 4.2. Components

There are different ways to implement a decentralized IMS. All approaches, however, will have to solve a similar set of problems, most of which have to deal with finding ways of ensuring trust in information without recourse to some authority. To get an idea of how this can work, the following components must be considered.

### 4.2.1. A Unique Identifier

To make a decentralized identity, a unique identifier that can be used in a decentralized manner has to be established. This is often referred to as Decentralized IDentifier (DID) [18]. Unlike traditional identifiers that are provided by the authority issuing the identity, DIDs are created by the entities themselves. DIDs are therefore independent of any centralized registry, identity provider, or a certificate authority.

A DID has a public and a private key; the former can be shared publicly with other entities, while the latter has to be kept for the DID owner; using this private key, the entity can digitally prove the ownership of a DID.

A DID is a permanent identifier in the sense that it never needs to change; it is also resolvable since it can be looked up to get its associated metadata. A DID is supposed to give an entity a lifetime encrypted private channel with another entity. It will not only be used for authentication, but also to exchange messages and VCs.

A person or an entity can create as many DIDs as needed for whatever purpose. Once a DID is created, its public part should be registered on a distributed ledger so that the actors involved in the relationship can look that DID up. The metadata associated with a DID is usually referred to as DID document; it simply contains information related to the DID and consequently important for the created relationship.

The syntax of a DID is the following as defined by the W3C: Scheme:Method:Method-Specific Identifier. The specification method defines how to read and write a DID and its document on a specific blockchain or distributed ledger [19]. Currently, active DID Method Specs are: Sovrin, Bitcoin Reference, Ethereum uPort, Blockstack, Veres One and IPFS. A DID method spec defines the syntax of the method-specific identifier, details any specific elements related to the DID document, and lastly describes the Create, Read, Update and Delete operations on DIDs and their documents.

### 4.2.2. DID Document

With a DID, a document is usually associated; it consists of a JSON-LD document giving additional information related to the DID. The purpose of this document is to describe keys' structure, the required protocols for authentication, and the service endpoints to interact with the identified entity.

More specifically, A DID document, as defined by the W3C, includes six components: the DID itself; cryptographic materials such as public keys used for authentication; cryptographic protocols to interact with the DID subject; the list of DID endpoints; auditing timestamps; a JSON-LD signature to verify the document integrity. A DID document can therefore be seen as the interface containing the required information to establish a communication channel with the DID owner.

### 4.2.3. Distributed Ledger

To make DIDs decentralized, a distributed ledger [20] has to be used. While blockchain is not required for decentralized identity, it can be a powerful distributed ledger for different aspects[4].

A blockchain provides a ready-made infrastructure for managing data in a decentralized but trustworthy way. Its promising features in terms of decentralization, immutability and transparency, easily allow the blockchain to be used as distributed register for DIDs.

Furthermore, by putting their hashes on the blockchain, entities' credentials can be notarized. By doing so, the blockchain acts as a timestamp and electronic seal. This undoubtedly provides both a proof of when the credential was created, as well as, an electronic seal that makes any tampering of the credential evident to any outside observer.

In addition, a blockchain can be a very efficient solution to record the access rights to information and the users' consents. Lastly, the blockchain may allow for process automation thanks to its smart contract execution. Indeed, in many real-world applications, verifiable credentials are used and verified to trigger a business workflow such as sending money to the user.

For all the above-mentioned reasons, this work focuses on the usage of the blockchain as a distributed ledger in the decentralized IMS.

[4] https://www.ibm.com/ blogs/blockchain/2018/06/self-sovereign-identity-why-blockchain/

### 4.2.4. Verifiable Credentials

Creating DIDs and their associated documents, and registering them in a distributed manner thanks to a blockchain will not suffice to enable the whole features of a decentralized identity infrastructure.

Another important building block – where most of the value is unlocked – lies in the usage of Verifiable Credentials (VCs) [21]. As previously defined, a VC is a cryptographically trustworthy piece of information about an entity's background.

It is usually shared as a trusted and verified proof that is linked to the distributed ledger by a public DID and credential definition written on-chain by the credential issuer. Typically, such trusted proof has the form of a digital signature, and can be verified using the public key of the issuer's DID. An example of a verifiable credential is a digitally issued certificate.

In a decentralized identity framework, VCs have to be transferred in a way that they are understandable and usable by any other system. Otherwise, VCs will have to be manually parsed; this will consequently prevent automated processes from being performed, and identities from being automatically transferred. To deal with this issue, standardization efforts of the schemes defining the structure and content of VCs should take place. JSON and some of its specialized versions is currently the most widely used standard for identity-related data.

### 4.2.5. Storage Agents

To enable their usage, VCs should be stored somewhere to make them available when needed. Additionally, private keys associated with DIDs must also be securely stored so that they are available to use when proving ownership. Storing such information is crucial for any decentralized identity system.

To store such private data, personal devices can be used such as a smartphone or a laptop, or some secure solutions provided by third parties. When solely under the control of the entity itself, identities and their associated data are considered as self-sovereign. Having data under the entity's control also makes it more inter-operable, allowing the employment of data on multiple platforms and for different purposes, and protecting the entity from being locked into one platform.

### 4.2.6. Private wallets

Finally, to implement a decentralized IMS, tools to manage DIDs, private keys and verifiable credentials will be required. Such tools are currently referred to as digital "wallets" [22]; they may come in the form of mobile phone apps, software, cloud or even hardware wallets. As with all other aspects of decentralized identity, the essential element here is that the wallet, and access to it, must remain under the entity's sole control.

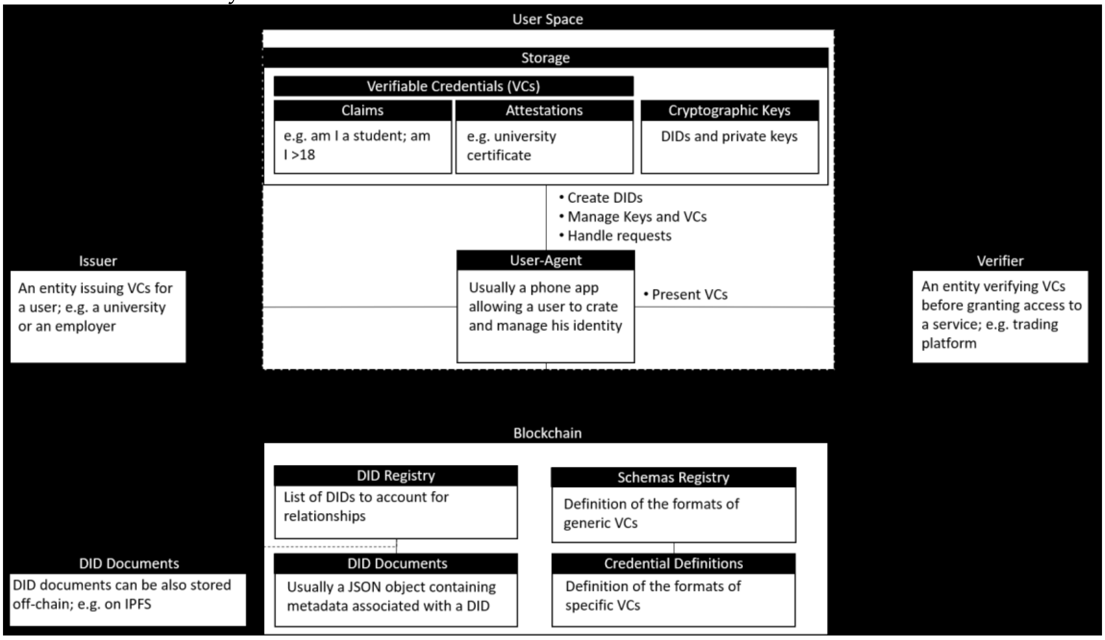
User Space
Storage
Verifiable Credentials (VCs)
Claims
Attestations
Cryptographic Keys
e.g. am I a student; am I >18
e.g. university certificate
DIDs and private keys
• Create DIDs
• Manage Keys and VCs
• Handle requests
Issuer
An entity issuing VCs for a user; e.g. a university or an employer
User-Agent
Usually a phone app allowing a user to crate and manage his identity
• Present VCs
Verifier
An entity verifying VCs before granting access to a service; e.g. trading platform
Blockchain
DID Registry
List of DIDs to account for relationships
Schemas Registry
Definition of the formats of generic VCs
DID Documents
DID documents can be also stored off-chain; e.g. on IPFS
DID Documents
Usually a JSON object containing metadata associated with a DID
Credential Definitions
Definition of the formats of specific VCs

**Figure 4:** Architecture of self-sovereign identity

### 4.3. Workflow

From a workflow point of view, as can be seen in Figure 5, a connection has first to be established between the user and service provider. This is initiated by the user as he/she is always the center of all actions in a decentralized identity context. Once the service provider is reached either via a website, mobile application or any other communication means, a DID has to communicated with the user. By doing so, the user can verify the identity of the service provider.

This latter has also to receive the DID of the user so that it can be verified that the service provider is communicating with the owner of the identity. Certainly, for the communication and verification to be made, the data related to the user agent's end points has to be retrieved from the DID document stored in the distributed ledger (i.e. the blockchain).

Once a secure communication channel between the user and service provider is established, any VCs can be sent, received and verified. That is, the service provider sends the user a request asking for some verifiable data (e.g. proof of address) in order to provide back the service.

The data model of the request is linear with the credential schema stored in the blockchain. Repeatedly, the user thanks to the communicator agent, will receive a notification asking for an approval of submitting the appropriate data in response to a service. As such, the user remains the only entity that can give a permission for the data to be submitted in their verifiable form.

From a service provider point of view, a notification will be received in response to the submitted demand. The response will contain the user's data either in an encrypted, predicate, or decrypted way. In all cases, the verification step is made in order to check that the received data is authentic, related to the user, has not been altered with and corresponds to what was required.

The most common way of exchanging data is to send the original attributes (e.g. age) encrypted with the public key of the service provider. This latter will later be able to decrypt the data and get its original value.

Obviously, that classical way may cause the leakage of some important data related to the identity of the user. An alternative form is to use Zero Knowledge Proof (ZKP) protocols whereby the user can prove to a service provider the possession of some information without revealing its original value. By doing so, the real values are kept protected while the service provider can still be able to verify some information about the user.

A last exchange model is to send encrypted data to the service provider. This latter will then use advanced techniques such as homomorphic computation to run their models over encrypted data. Although that technique maximizes the user's privacy, its application is still not widely adopted due to its high computation time [23].

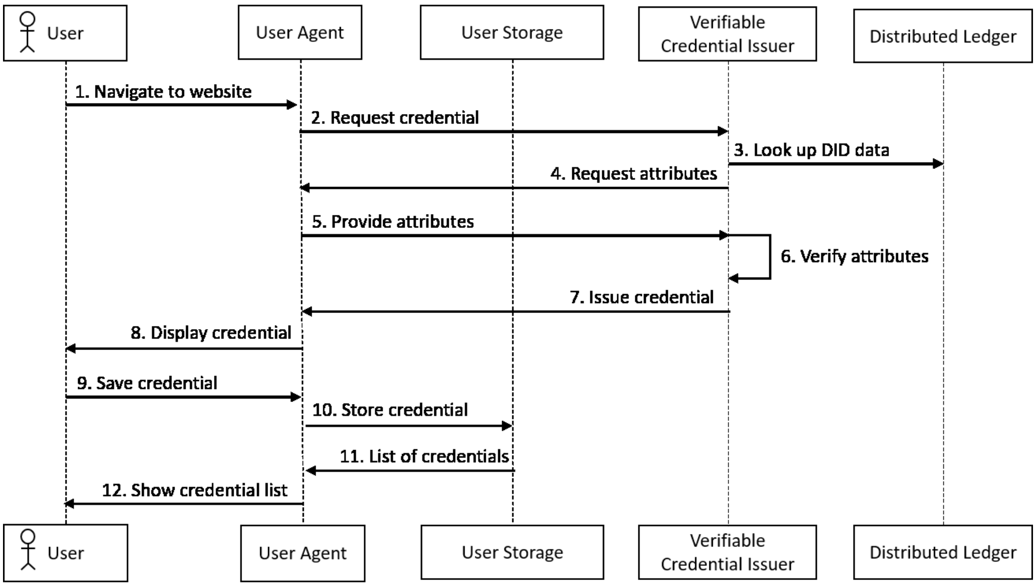

**Figure 5:** Credential workflow

### 4.4. Evaluation Criteria

In order to generalize the SSI, its principles have been formalized as discussed in [24]. Those principles attempt to ensure the user control in an identity system that balances transparency, privacy and fairness. In this survey, the analysis of the existing solutions in subsection E will be based on those principles; a summary table of the comparison is shown in Table II; the ten principles can be summarized as follows:

- **Existence:** Users must never be wholly digital; an independent existence must exist. An SSI must therefore be based on a physical identity.
- **Control:** Users must fully control their identity, refer to it, update it, or even hide it.
- **Access:** Users must have access to their data and be able to retrieve all their own claims. There must not be any personal data hidden from its owner.
- **Transparency:** All the SSI components must be transparent. An SSI solution must be open on how it works, how it is managed and updated. The used algorithms must be open-source, free, and as independent as possible of any particular organization or architecture.
- **Persistence:** An identity must be long-lived; it can only be removed by its owner. A claim associated with that identity can be updated or removed, but the identity must be long-lived.
- **Portability:** The identity's attributes, claims and services must be transportable by their owner. An identity must never be solely managed by a third party since that later can disappear.
- **Interoperability:** An identity must be as widely usable as possible. A true SSI is globally adoptable and must not only be limited to certain niches.
- **Consent:** Users must freely agree on how their identity attributes and data are exploited. Identity's information must not be shared without having a consent from the user.
- **Minimalization:** Disclosing identity attributes must be minimized. Only the necessary piece of information must be shared. An example is to only share whether someone is older than 18 instead of revealing the full date of birth.
- **Protection:** The users' rights against the powerful must be protected. The users' rights and freedoms are more prioritized than the network's needs.
- **Provable:** Identity claims must be able to be verified, for example by trusted third parties.

### 4.5. Existing Solutions: Description and Analysis

After explaining the main components and the generic architecture of an SSI system, the most relevant SSI solutions are overviewed, discussed, analyzed and assessed based on the previously defined ten evaluation criteria.

#### 4.5.1. uPort

uPort [25] is an open-source solution based on Ethereum. It aims at providing decentralized identity for all. An identity is first created by the user through a dedicated mobile application that stores all user's identity data, including the private keys used to sign and share claims.

Once the user creates an identity, two smart contracts "controller" and "proxy" are automatically deployed on the Ethereum blockchain.

The proxy contract has a reference toward the address of the controller contract; the proxy functions can only be invoked by the controller; the address of the proxy comprises the users' unique uPort identifier (uPortID); a user can create multiple unlinkable uPort identifiers.

Mapping identity attributes to a particular uPortID is done by deploying a "registry smart contract". This registry can be queried by any entity, however, only its owner can update its attributes; since the blockchain is not devoted to store large amount of data, the JSON attribute structure is hashed then stored in the registry; an external secured distributed file system such as IPFS is used to store the attributes themselves.

To deal with the theft of the user's mobile device or keys loss, uPort uses a social recovery protocol whereby the user nominates a set of trustees who can vote to replace the public key residing in the controller contract. Once those trustees reach a voting quorum on the new public key, the controller updates the lost key with the newly generated one.

That recovery protocol allows the user to have a persistent identifier even after losing the associated keys. An uPort user is capable of signing claims about other entities and publishing them on the ledger so that the trustworthiness of such attestations is increased.

The main limitation of uPort lies in the lack of portability. That is because only other uPort identities are able to attest. Additionally, the issue of interoperability may arise since it is primarily based on Ethereum.

Furthermore, while the social recovery process allows recovering the ownership of a compromised or lost uPortID, the set of trustees can be an input of attack if they decide to collude against the user. If an uPort application is compromised and the list of trustees is maliciously changed, the uPortID will be permanently compromised.

Since uPort is primarily based on smart contracts, a malicious blockchain node may trace and link all activities related to one uPortID; as such, the real identity of the user, as well as, all the associated actions may be known. This may certainly comprise the privacy of users.

### 4.5.2. Sovrin

Sovrin[5] is an open-source decentralized identity network built on top of a permissioned distributed ledger. Sovrin is public in the sense that anyone can send read transactions, however, only trusted institutions such as governments, banks, universities, etc. can have nodes that participate in the consensus protocol; the writing access is thus permissioned.

A mobile application and an agent are used to allow for interactions between the user and the rest of the network. These components also help users to manage cryptographic keys.

As in uPort, Sovrin uses a key recovery protocol based on the user's nomination of trustees. After a request from the user, the trustees must reach a quorum in order to send an identity record change transaction, which must be validated by stewards.

With sovrin, no claim will be registered in the blockchain. Moreover, it defines also an agent to agent communication without relying on the shared ledger that provides more privacy and confidentiality.

The main issue related to Sovrin is the presence of predefined institutions as middlewares between the user and the DL; although their role is ultimately to assess the identity of the user, such position may allow them to access important information related to the users' identities.

As such, the consent and protection properties of the identity principles may be compromised. Additionally, Sovrin does not provide any guarantee regarding the correct functioning of the network agents. Therefore, the provable property may also be compromised.

### 4.5.3. ShoCard

ShoCard[6] is an IMS that leverages blockchain and cryptographic hashes to bind together a user identifier, an existing trusted credential such as passport or driver license, and additional identity data. Thanks to ShoCard mobile application, users first scan their identity credentials; such data is stored in an encrypted manner within the user's mobile device; the data is also written on the Bitcoin ledger to enable a-posteriori data validation; the identifier of the resulting Bitcoin transaction is the user ShoCardID and is stored in the mobile device as a pointer toward the ShoCard seal. In order to associate certificates to a ShoCardID, a user interacts with an identity provider through a process called certification.

Since the user needs to provide some attributes to relying parties such as identity providers and may not want to lose them if the mobile device is lost, a ShoCard server may be used to store encrypted version of the attributes.

Although ShoCard promotes for decentralization, the existence of an intermediary central server managing the distribution of encrypted certificates between ShoCard users and relying parties may open the door for data breaches.

[5] https://sovrin.org/

[6] https://shocard.com/

In addition, the ShoCard role as a middleware may create uncertainty about the long-liveness of a ShoCardID; if the company disappears, ShoCard users would be unable to use the system with their already acquired certifications.

ShoCard can then be more centralized in practice than its reliance on a Distributed Ledger. Besides, as previously mentioned, for each certification the user is required to provide a trusted credential, users are required to embed more personal information than the necessary. Adding to all that, the dependency on bitcoin may cause additional issues such as cost and transactions validation time.

### 4.5.4. IDchainZ

IDchainZ[7] is an extension that comes in the form of a proof of concept of a smart ledger called ChainZy. IDchainZ allows users to have a key ring of trusted and certified identity claims and enables relying parties to exchange know-your-customer and anti-money laundering data. Those exchanges can be practically extended to all types of documents.

As a first step, the user provides the certifier with documents for certification; the certifier validates the documents and creates IDchainZ identity; the certifier also creates a master ring, sends its key to the user, and uploads the certified documents to IDchainZ ledger; once the user receives the key, the certifier removes its copy of key and the underlying documents.

Once a service provider asks for the document, the user creates a subkey from his master one and sends it to the requester; with this subkey, the duration, shared information, number of accesses and use limits are associated and controlled; that allows the user to fully control the lifecycle of the identity attributes.

Assessing the compliance of IDchainZ with respect to the ten identity principles is difficult since not much details about the solution are given. Overall, the solution highly depends on IDchainZ itself; thus, identities may not be persistent, as well as, inter-operable with other platforms. When it comes to the minimalization of data, proves are not given.

### 4.5.5. EverID

EverID[8] is an SSI and value transfer system based on blockchain; it is used to store identity attributes, certifications and bio metrics. EverID facilitates the users' identity verification process and enables a secure transfer of value between the network's members. It also uses smart contracts to provide personal data ownership whereby the user controls how the identity attributes are shared, for how long, with whom and for which purpose.

The user in EverID system does not need to have a mobile device since the government ID, biometrics data and third-party certifications, which constitute the DI, can be stored in a cloud. Nonetheless, the minimalization principle in EverID is not fully respected. When a piece of identity information is required to verify a claim, the user has no choice but fully disclosing the full data. EverID is also not open source, which compromises the provability and transparency principles of an SSI.

### 4.5.6. LifeID

LifeID[9] is an open source platform that enables independent creation and management of users' DI. Users control and approve all the online real-world transactions, without relying on third parties. Data in LifeID is stored on users' devices and only the required piece of information is disclosed when needed; this is enabled thanks to ZKP protocols.

LifeID offers three options to back up and recover identity attributes: cold storage, a list of trustees and selecting a trusted organization. This gives users more solutions to fight against data breaches and device theft.

LifeID also provides an open-source software development kit (SDK) allowing ecosystem partners to use LifeID as an identity platform layer. An SDK has been also developed to facilitate

[7] https://www.chainzy.com/

[8] https://everest.org/

[9] https://lifeid.io/

the creation of a broad variety of identity solutions across a wide range of devices; this certainly maximizes the inseparability of the solution.

LifeID also reduces the implementation friction since it has been designed to be run on any blockchain supporting smart contracts. Although the LifeId white paper is very rich and it claims meeting all the identity principles, unfortunately, technical details are not given; additionally, the solution architecture itself is not given.

Analyzing the GitHub repository activities shows that the project is not active. This certainly does not fulfill the transparency and provability criteria of SSI.

#### 4.5.7. SelfKey

SelfKey[10] is an SSI network whereby users' data is stored on a personal device. The user can also back up identity information onto another device or a personal backup solution. When a third party asks for collecting specific information, the user is first notified and a consent is demanded before the information is revealed. When the collection is approved, ZKP protocols are used to ensure that only the minimum necessary amount of data is disclosed.

In SelfKey, users' identity claims can only be verified by trusted entities to ensure the provable principle of the SSI. Despite the solid architecture of SelfKey, relying on a private instance of Ethereum with a Proof of Authority (PoA) consensus opens the door for many privacy issues.

First of all, it is known that PoA is not a byzantine fault tolerant algorithm [26]; this means that one malicious participant may compromise the whole consensus.

In addition, in such consortium context, predefined actors may be considered as third parties between the user and the DL; this may lead to anonymity, traceability and linkability issues. For such reasons, SelfKey compromises the persistence, consent and protection principles.

#### 4.5.8. Civic

Civic[11] is a blockchain based identity solution. It provides three major features:

- **Security Identity Platform:** It's a decentralized identity verification platform. Users enroll via an application. Their data are locally stored on the device and protected using biometric authentication. A scan of a legal document (passport, ID Card, driving license) is mandatory to verify the identity, and the picture on the document is compared to a selfie to attest that the user is the real owner of the document. Public identifier, hashed private data and validity of the identity is published into a public blockchain.
- **Reusable KYC:** It permits to attest to a service provider that an identity has been verified by Civic. A proof of verified identity is stored in Civic blockchain. This service also manages the authorization to access the user personal data; no access is granted without the user consent.
- **ID Codes:** are used as supports to enable using the identity without entering a passphrase. ID Codes are provided by service providers that want to access the identity; they are associated with the specific attributes needed by the service. ID Codes can be verified via the blockchain to avoid malicious service providers accessing the identity's data.

All private data and cryptographic keys are stored locally inside the user's device, and can be backed up to a personal account on a cloud based or distributed storage platform. In Civic white paper no information is provided about whether the platform is open source or not. That may cause a lack of transparency for the solution.

Furthermore, the approach promotes for data minimization, but technical details are not given making the provability of the solution hard to assess. Finally, the solution is based on the existence of the Civic entity either for validating users' identities, or to design the business workflow through CVC tokens. Users' identities may therefore be dependent on the existence of Civic itself; therefore, they are not persistent. Such dependency may also cause interoperability issues.

[10] https://selfkey.org/

[11] https://www.civic.com/

### 4.5.9. THEKEy

THEKEy[12] is a blockchain based solution for DI running on the NEO blockchain. Biometric data are used in the first place to validate and identify later users.

During the registration phase, biometric data are verified face to face or compared with existing biometric data issued from another registration. Users have full access to their data but the storage of data is not decentralized.

When a service provider asks for accessing the user identity, a notification is sent to the user app; the user then gives a consent thanks to biometric signature; later, THEKEy platform provides the required data, and checks if the required attributes matches the service's needs.

The workflow of this approach indicates that the user identity will highly depend on a third party "THEKEy" platform. As such, no guarantee can be given to ensure the user's privacy and data security, as well as identity's persistence and portability. In addition, assessing the data minimization, nothing proves that only the required data will be provided for service providers except the validation of the platform.

### 4.5.10. Bitnation

Bitnation [27], or cryptonation, aims to create a decentralized nation by recording vital records, identity and other legal events between citizens (e.g. marriage contracts, birth certificates, companies' registration, lands title, etc.) using blockchain technology.

Bitnation implements a "Decentralized Border-less Voluntary Nation" using a blockchain-agnostic network called the "Pangea Juridiction", which uses the Ethereum blockchain, IPFS technology and secure scuttlebutt protocol.

Bitnation proposes a mobile application that allows users to (1) create an account, (2) create, join or quit a Nation, (3) list available smart contracts and (4) start a chat. Bitnation allows users to independently create and control their own identity.

In addition, it gives a permanent access to an identity and creates an environment where users can continue to build up their portable identities thanks to smart contracts that are executed on the Ethereum blockchain. Bitnation does not provide data minimization nor verified credentials. Additionally, it does not deal with standardization making the identity not inter-operable with other solutions.

## 4.6. Evaluation

As a conclusion of the above description, Tables I and II summarize the assessment of all the detailed solutions regarding the ten principles of SSI presented in the subsection IV-D. Additional information is also provided such as the year the solution was created, the existence of a white paper describing the solution, the availability of the source code, the distributed ledger used in each solution and lastly the blockchain type.

**Table 1.** Non-functional assessment of SSI solutions

| | **Creation Year** | **White Paper** | **Source Code** | **Distributed Ledger** | **Blockchain Type** |
|---|---|---|---|---|---|
| Uport | 2016 | Footnote[13] | Footnote[14] | Ethereum | permissionless |
| Sovrin | 2016 | [28] | Footnote[15] | Indy | permissioned |
| ShoCard | 2015 | [29] | Footnote[16] | Bitcoin | permission(less,ed) |
| IDchainZ | 2014 | - | - | - | - |
| EverID | 2016 | Footnote[17] | - | Ethereum | permissioned |
| LifeID | 2017 | Footnote[18] | Footnote[19] | LIFEID | permissionless |

[12] https://www.thekey.vip

[13] https://blockchainlab.com/pdf/uPort_whitepaper_DRAFT20161020.pdf

[14] https://github.com/uport-project/

[15] https://github.com/sovrin-foundation/

[16] https://github.com/shocardinc

[17] https://coinosophy.files.wordpress.com/2018/05/everid-whitepaper.pdf

[18] https://lifeid.io/whitepaper.pdf

[19] https://github.com/lifeID

| | | | | | |
|---|---|---|---|---|---|
| SelfKey | 2017 | Footnote [20] | Footnote [21] | Ethereum | permissionless |
| Civic | 2015 | Footnote [22] | Footnote [23] | Ethereum | - |
| TheKey | 2014 | Footnote [24] | - | NEO | - |
| Bitnation | 2014 | Footnote [25] | Footnote [26] | NEO | - |

**Table 2.** Functional assessment of SSI solutions

| | Cont | Acce | Trans | Pers | Port | Inte | Cons | Mini | Prot | Prov |
|---|---|---|---|---|---|---|---|---|---|---|
| Uport | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Sovrin | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 |
| ShoCard | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |
| IDchainZ | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| EverID | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 |
| LifeID | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| SelfKey | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 |
| Civic | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| TheKey | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| Bitnation | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 |

Cont: Control | Acce: Access | Trans: Transparency | Pers: Persistence | Port: Portability |
Inte: Interoperability | Cons: Consent | Mini: Minimalization | Prot: Protection | Prov: Provable

### 4.7. Challenges and Discussions

This section addresses the challenges hindering the development and build of a complete decentralized IMS. As shown in Figure 6, we have divided those challenges into two main categories; technical and non-technical ones.

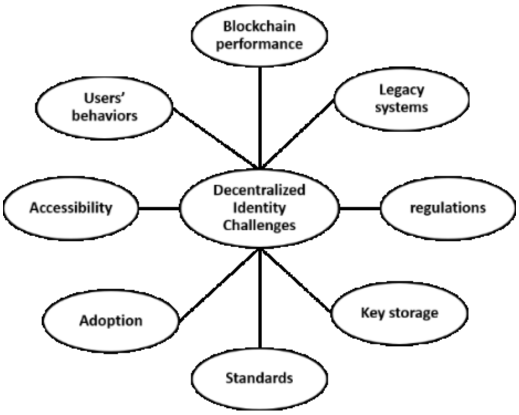

**Figure 6:** Decentralized Identity challenges

#### 4.7.1. Technical Limitations

Many decentralized identity believers claim that nowadays all the tools and techniques are present to build an efficient system whereby identities' attributes and their associated data are fully under the control of the end user.

Those claims have been recently reinforced with the rise of the blockchain technologies that are promised to enable DIDs and VCs. Those believers went much further by explicitly indicating that all the components of the decentralized IMS can be currently built without any limitation and all the user's data privacy and control's requirements can be fulfilled.

This section ultimately aims at pointing out some technical points that are still not resolved, and consequently will slow down the decentralized identity shift. Those limitations are summarized as follows:

[20] https://selfkey.org/wp-content/uploads/2019/03/selfkey-whitepaper-en.pdf

[21] https://github.com/SelfKeyFoundation

[22] https://tokensale.civic.com/ CivicTokenSaleWhitePaper.pdf

[23] https://github.com/civicteam

[24] https://www.thekey.vip/#/WhitepaperDownload

[25] https://tse.bitnation.co/documents/

[26] https://github.com/Bit-Nation

#### 4.7.1.1. Usage of the Blockchain as a Distributed Ledger

As already mentioned in previous paragraphs, the blokchchain is nowadays considered as the most efficient technology that can be used as a distributed ledger for a decentralized IMS.

Such technology is primarily used to store DIDs; its usage can be extended further to store DID documents and users' consents, notarize credentials, and automate business workflows thanks to smart contracts. Analyzing a decentralized IMS cannot therefore be fully accomplished without studying the technical limitations of the blockchain itself.

To start with, the anonymity issue in the blockchain is discussed. Indeed, in order to register a DID in the blockchain, the user must send a transaction; to do so, the user usually uses a pseudonym that is associated with a blockchain address. As long as there is no link between that address and the user's physical identity, the sent transaction remains anonymous. But as soon as any link is established between the blockchain address and who the user is — whether it be an IP, an email, or anything identifiable — then the cover is blown.

Having said that, it can be noticed that all the DIDs created by a specific user are prone to be known for a blockchain tracer. Knowing this does not consequently mean that all the user's relationships are traced.

Indeed, if the relation between two DIDs (e.g. a user's DID and a service provider DID) is also put on a public ledger, tracing and linking all user's activities would then be possible. On the contrary, if the relationship itself is made privately (i.e. in a secure connection between only the actors involved in the relationship), a blockchain tracer cannot know the purpose of the user's DIDs. That is similar to saying that the blockchain can help in the build of a decentralized IMS, however, this latter does not help a blockchain in solving its identity issues.

One cannot talk about the usage of a blockchain without mentioning its storage limitation. Imagining that all the DIDs resulting from establishing relationships between actors (e.g. users, service providers, connected devices) will be stored on a distributed ledger. This will obviously be an important problem to deal with even if multiple blockchains are used to store DIDs.

Furthermore, adding DIDs to a blockchain will not be in real time. That is, whenever a user expresses a willingness to create a relationship with another entity, a DID registration transaction should be sent to the blockchain and its validation should be waited for. This will undoubtedly make the usage of any IMS dependent on the transaction validation time. Currently, the throughput of the most secure blockchain technology (i.e. bitcoin) does not reach even 10 transactions per second.

Moreover, in most public blockchains, for a transaction to be validated and eventually written on the ledger, some fees are required to be sent with. Obviously, when the user sends transactions to create his/her DIDs, the cost question emerges. This is not only related to the cost value itself, but also to the establishment of the payment mechanisms themselves.

To deal with the scalability issue of public blockchain technologies, many people suggest the usage of consortium ones [30]. While this will certainly solve the problem, subsequent problems will arise. Indeed, a permissioned blockchain assumes that a set of actors are already set together to collectively run a blockchain.

As such, the consortium is highly dependent on the initial created members. This implies, that the validation of the DIDs is dependent on a predefined set of actors, which may go against the decentralization idea of the identity.

#### **4.7.1.2.** Key **Storage**

The key storage is another technical point that should be considered while developing a decentralized IMS. When creating DIDs, their associated private keys and VCs have to be securely stored somewhere in order to be used when needed.

Currently, most of the solutions consist of storing such precious private data in user's devices such as a smart phone, personal computer or even a hardware device. The legitimate question of what will happen in case of losing such storage tools is still to be addressed. Clearly, losing a storage tool implies losing the identity of its owner.

To deal with such issue, many experts claim that the user is supposed to have multiple devices such as a phone and a personal computer at the same time. Therefore, whenever a device lost occurs, the user can use one of his other devices to revoke or rebuild his identity.

Obviously, this does not completely solve the problem since a user may lost all of storage devices. More advanced solutions suggest having a backup copy of the user's identity private data on a cloud storage managed by a third party [31]. These backup data can be encrypted by a password only the user knows, or by more advanced techniques such as the user biometric data. Although this may solve the problem, but the security of such approaches is still questionable.

Indeed, storing sensitive data within a cloud storage third party is not recommended even if the data is encrypted or hashed. Giving encrypted data to someone refers to giving an infinite time to break the encryption and get the real value behind the data. This becomes very dangerous when data refer to personal attributes.

#### 4.7.2. Non-technical Issues

In addition to the above described technical limitations, the development of decentralized IMS(s) is facing non-technical issues related to the identity legal systems, as well as, the acceptance degree associated with such important shift. In this section, those challenges are discussed.

##### 4.7.2.1. Legacy Systems

To allow a user building an identity, credentials should be collected and signed by trust anchors, such as banks and government agencies. Nowadays, many of such trust anchors are running centralized identity systems where they themselves store/manage the users' identity attributes. Obviously, changing the nature of relationships, as well as the ownership of users' valuable identity data will not be an easy task. In addition, not only trust anchors should adapt their infrastructures, but also service providers and other relying parties.

##### 4.7.2.2. Regulations

In order to benefit from the advantageous of decentralized IMSs, technologies, standards and regulatory frameworks have to evolve. Currently, operating models are designed for centralized systems. Making the decentralized shift will therefore take an important time before it becomes widespread. Additionally, assigning liability for potential breaches or private data abuses may be especially very complex. Individuals may need education to adopt decentralized systems and responsibly use them.

##### 4.7.2.3. Standards

To unleash all the potential of decentralized identities, standards should be developed so that DIDs and their documents, as well as, credentials can be easily read and verified. Otherwise, siloed systems will be created where each service provider has its own format of DIDs and credentials, and interoperability between solutions will be absent or difficult to be established. This standardization process will certainly take some time, and require a lot of coordination between the digital actors across the planet.

##### 4.7.2.4. Adoptions

As the decentralized paradigm represents an important shift in the management of digital identities, its adoption will primarily depend on the adoption level of its actors. On the business side, actors providing services that are built on top of identity solutions need to reorganize their business models in order to benefit from the potential of this shift such as reusable credentials and verified user data. On the user side, the challenge is more complex.

The user's adoption of a new technology may depend on many factors such as their knowledge, the easiness of the technology, the usage cost, the willingness, etc. It is clear that a decentralized IMS will face the challenge of complexity; breaking the status quo of convenient centrally-managed identity where interactions are routed through a third party will be very hard.

### 4.7.2.5. Accessibility

The accessibility of the decentralized identity is another point that can by no means be disregarded or underestimated. That particularly refers to ensuring that everyone having the right to have an identifier or credential will be able to use them in the new identity architecture.

That comprises disadvantaged or vulnerable population, as well as, people with physical or mental disabilities, children, the elderly, and those without access to technologies. Those categories cannot therefore be excluded or precluded while building any decentralized IMS.

The accessibility also raises issues related to the technical infrastructures required to hold an identity. Smart phones and personal computers will probably be a must for an individual to manage a decentralized identity, however, many people are nowadays still with no access to such tools. In addition, many people will require additional means to control their identity due to physical or mental disabilities. This will also mean that specific measures including laws, regulations should take place in order to cope with such cases. Clearly, such cases that are crucial for any IMS will make the build of a decentralized one more challenging.

### 4.7.2.6. The Behaviors of Actors

Users' behaviors must be taken into account during the development of any IMS since it may highly affect the adoption of the solution. The most common examples are solutions with too complicated workflows.

Users will just not adopt the solution and even if they are forced to use it. They will try instead to find easier alternatives often at the expense of privacy or security. Users' concerns about the usage of their private data is growing due to the amount of data leakage but there is still a vast majority trusting companies in how they are exploiting those data.

Password based identification is another example of identification solution which is weakened by users' behavior. Usage of weak password or repeated usage of the same password is common for a large majority of users. Another problem with password-based solutions is the loss of the password. If the service doesn't provide recovery mechanisms, it will not be adopted by a large part of users.

Authentication using biometrical information are new identification alternatives, which also decrypt local stored private information without the possibility for a user to weaken the solution's security. Due to the complexity of decentralized identity, users therefore may adopt alternative solutions where third parties manage their identities. By doing so, the traditional identity paradigms will be reproduced.

When asking for the minimal data to enable a service, businesses may show an undesirable behavior where they ask for more information than the necessary. Clearly, such action is very difficult to be detected in real time.

## 5. Related Works

In recent studies, several solutions exist where a blockchain or any other distributed ledger is integrated in order to design a decentralized IMS in different application scenarios. Survey articles have attempted to review these proposed solutions in varied degrees of depth and scopes. In the following, we discuss a comparative table between the existing surveys based on seven criteria:

**Identity Definition:** This criterion checks whether the survey paper proposes/details an identity definition or not.

**Identity Components:** some works define a generic architecture for the SSI and detail its components. This criterion aims at checking whether the identity architecture and its components are discussed or not.

**Use Cases:** identity solutions can be integrated into different use cases. For each survey, we study if this integration was detailed or not and for which use cases.

**Challenges and Future Works:** an interesting part of a survey is the proposed challenges and future works. In the studied surveys, some papers completely omit the challenges of the novel IMSs, others present the challenges with details, and another class where challenges are presented without details and discussions.

**Life Cycle:** when to deal with identity and credentials, an important task is to detail the different steps related to their life cycle such as creation, issuance, recovery, suspension, revocation, discoverability and transferability. Therefore, this criterion checks for each of the analyzed surveys whether the life cycle of the identity is detailed or not.

**Evaluation:** another considered criterion is the comparison of the different surveys regarding the number of studied solutions and the proposed criteria to compare them with respect to the SSI principles.

**Year of Publication:** the final criterion is related to the publication year of the survey. That indicator is important since the SSI concept has seen enormous evolution in the last years; it is important thus for a survey to cover the latest status.

As shown in the Table III, nine surveys were studied. In [32], the concept of the self-sovereign identity and the challenges and opportunities are discussed in an informal way. In [33], authors aim to provide a formal and rigorous treatment of the concept of self-sovereign identity using a mathematical model. It also presents the essential life-cycles of an IMS with detailed sequence diagrams. In other work [34], the identity was defined based on different use cases. The authors try to study a large set of IMSs with different levels of maturity and availability.

This study is based on three sets of criteria: compliance and liability, user experience and technology criteria. Their work discusses the integration of an IMS in different use cases. However, it will be more relevant to filter the set of approaches in order to eliminate those in the initial phase or without implementation.

Moreover, the comparison is based on a high-level set of criteria. Besides, the work did not detail the challenges and future works. Therefore, the provided comparison may be recommended for no technical readers. Very rich studies are made for specific use cases in [35] and [36]. The former survey focuses on IMSs for IoT, while the latter's attention is given to discussing the identity paradigm across different political and economic domains.

In [37], [38], a short and simple definition of the identity is given. More presentations of the components of an SSI are detailed in [38]. However, authors do not give much attention to the comparison of the selected solutions. In [39], a new and relevant study has been proposed. The main output of this work helps readers to understand the terminologies and standards of IMS(s). Moreover, they provide an adequate amount of details related to the security, presentation and data sharing challenges.

Finally, in [40], authors interestingly provided a rich comparison of SSI solutions regarding the usage of the blockchain. Two main categories of solutions have been detailed: those based on the blockchain technology as a distributed ledger and others that use classical distributed solutions.

In the following, the main contributions and differences of this paper compared to the existing surveys are presented:

- In this paper, a global and generic architecture for an SSI system is proposed along with a detailed presentation of its mains components, principles and lifecycle. The ultimate objective is to provide a generic overview in this field of research. Unlike other surveys, the explained architecture and components in this paper are highly abstract and independent of a specific use case.
- A complete strategy has been established in order to assess various SSI solutions. That strategy highly considers the community feedback on the solution itself as well as the maturity of the source code and the inter-operability with other solutions. The availability of the documentation and tutorials explaining the usage and deployment of the solution are also analyzed. Since the SSI technology has seen enormous evolution, this paper focuses on the latest version of the studied solutions. More importantly, technical and non-technical limitations that hinder the complete adoption of the new SSI paradigm have been presented and discussed in this paper. Some recommendations and research directions have been also given so that the SSI technology can become more secure, scalable and robust.

**Table 3.** Comparison of existing identity management surveys

| Paper/ criteria | Identity definition | Identity components | Use cases | Challenges/ future works | Life cycle | Evaluation | Year of publication |
|---|---|---|---|---|---|---|---|
| [32] | Informal definition | Not detailed | Not detailed | It presents some challenges without detail | Not detailed | Not detailed | 2017 |
| [33] | Mathematical definition with a taxonomy proposition | Not detailed | Not detailed | Not detailed | Different sequence diagrams are detailed | 4 solutions | 2019 |
| [34] | Defined based on different use cases | Not detailed | Well detailed | It presents some challenges without details | Not detailed | 43 solutions | 2019 |
| [35] | Formal definition | Not detailed | detailed for IoT | Detailed for IoT use cases | Not detailed | 15 solutions | 2018 |
| [36] | The Origin of the identity and its evolution are detailed | Not detailed | Identity across political and economic domains | Not detailed | Not detailed | Not detailed | 2018 |
| [37] | A hierarchical model is presented | Not detailed | Not detailed | Not detailed | Not detailed | 3 solutions | 2019 |
| [40] | A short definition | Not detailed | Not detailed | A brief description of some challenges | Not detailed | Comparison with blockchain and Non blockchain solutions | 2019 |
| [38] | A short definition | High level overview of the SSI architecture and its components | Not detailed | Not detailed | Not detailed | Not detailed | 2018 |
| [39] | A terminology, list of standards, and fundamental building blocks are presented | Detailed | Well detailed | Detailed challenges related to the security issues, presentation and data sharing | Detailed | Not detailed | 2019 |

## 6. Conclusions

This paper addressed the topic of digital identity. A definition was first given; the current digital area was then discussed from an importance, opportunity and risk points of view. Later, two classical identity paradigms were presented, namely the centralized and federated identities; their schemes were explained and their strengths and limitations were discussed.

The focus along this paper went then to the decentralized identity that is the novel form of digital identity; its definition, importance and workflow were discussed. In this context, the self-sovereign identity was introduced; this new identity paradigm is a more advanced form of the decentralized identity where not only the identity attributes are remained under the control of the user, but also the data and their associated actions. The principles and components of such paradigm were also explained. In order to implement an SSI system, the most important components that have to be implemented were highlighted.

To answer the question whether having an SSI system is currently feasible, one section was devoted to discuss the main challenges and limitations. It was argued that not only business and social factors are still hindering the build of a complete SSI system, but there are also some technical issues. To complete this literature review, a set of existing solutions claiming the development of SSI systems were overviewed. Their designs were described, as well as, their capacity to cope with the ten SSI principles. Lastly, a literature review was made whereby existing surveys were detailed and compared according to several criteria.

The main conclusion resulting from this work is that building a decentralized identity system that efficiently manages users' identities while protecting their data and actions is not yet fully feasible. This is due to technical and non-technical factors. From a technical point of view, focusing on the scalability and inter-operability topics of the blockchain, as well as, the identity issues that may arise in the blockchain itself such as link-ability, trace-ability and anonymity of data is recommended. Besides, efficiently handling users' keys is a must for any SSI system. Till now, this key management topic is not fully dealt with.

From non-technical points of view, producing standards for the decentralized identity components is highly important. This work has been already started by some organizations and consortiums, such as Sovrin, the identity foundation, European forum, etc., and it should be extended more to reach a broader range of private and public organizations.

Although this paper provided descriptions and qualitative analysis of IMSs including the latest SSI one, there is still a seamless need to include more quantitative studies in that regard. To fill that gap, assessing the scalability of a blockchain-based SSI solution will be a part of the future works. More precisely, the throughput and latency of the Sovrin IMS will be evaluated against the active number of users. Moreover, statistical and business analyses will be done regarding both the adoption cost and the acquisition rate of the new SSI paradigm.

Finally, an equal importance will be given to quantitively evaluating the security levels of SSI solutions. That has certainly become a must, given the rise of quantum computing and post-quantum cryptography.

## References

[1] Li, Y., Yu, W., Li, X. and Yang, Z., 2020. Research on the evolution of global internet network interconnection relationship in 21 years. *China Communications*, *17*(8), pp.158-167.

[2] Catlin, T. and Lorenz, J.T., 2017. Digital disruption in insurance: Cutting through the noise. *Digit. McKinsey*.

[3] Laurent, M., Denouël, J., Levallois-Barth, C. and Waelbroeck, P., 2015. Digital identity. In *Digital identity management* (pp. 1-45). Elsevier.

[4] Windley, P.J., 2003. Understanding digital identity management. *The Windley Group*.

[5] Labong, R.C., 2019. Identity Theft Protection Strategies: A Literature Review. *Journal of Academic Research*, *4*(2), pp.1-12.

[6] Hammouchi, H., Cherqi, O., Mezzour, G., Ghogho, M. and El Koutbi, M., 2019. Digging deeper into data breaches: An exploratory data analysis of hacking breaches over time. *Procedia Computer Science*, *151*, pp. 1004-1009.

[7] Goel, R.K., 2019. Identity theft in the internet age: Evidence from the US states. *Managerial and Decision Economics*, *40*(2), pp.169-175.

[8] Moriarty, K.M., 2020. Authentication and Authorization. In *Transforming Information Security*. Emerald Publishing Limited.

[9] Khalil, M.M., Lamison, M.R. and Dubagunta, S., Verizon Patent and Licensing Inc, 2020. *Identity management via a centralized identity management server device*. U.S. Patent Application 15/929,806.

[10] Kumar M, N. and Honnavalli, P.B., 2020. Dynamic Federation in Federated Identity Management. *Suganthi and Honnavalli, Prasad B, Dynamic Federation in Federated Identity Management*.

[11] Kubach, M., Schunck, C.H., Sellung, R. and Roßnagel, H., 2020. Self-sovereign and Decentralized identity as the future of identity management?. *Open Identity Summit 2020*.

[12] Zhu, X. and Badr, Y., 2018. Identity management systems for the internet of things: a survey towards blockchain solutions. *Sensors*, *18*(12), pp. 4215.

[13] Pöhn, D. and Hommel, W., 2020, August. An overview of limitations and approaches in identity management. In *Proceedings of the 15th International Conference on Availability, Reliability and Security*, pp. 1-10.

[14] Amoli, G., Kala, M. and Chaurasia, J., 2019. Comprehensive Security Analysis of Federated Identity Management. *Journal of Communication Engineering & Systems*, *7*(1), pp.11-16.

[15] Lin, X., 2020. *New Innovations in eIDAS-compliant Trust Services: Blockchain* (Bachelor's thesis, Universitat Politècnica de Catalunya).

[16] Sadqi, Y., Belfaik, Y. and Safi, S., 2020, March. Web OAuth-based SSO Systems Security. In *Proceedings of the 3rd International Conference on Networking, Information Systems & Security*, pp. 1-7.

[17] Kubach, M., Schunck, C.H., Sellung, R. and Roßnagel, H., 2020. Self-sovereign and Decentralized identity as the future of identity management?, *Open Identity Summit 2020*.

[18] Alzahrani, B., 2020. An Information-Centric Networking based Registry for Decentralized Identifiers and Verifiable Credentials. *IEEE Access*, *8*, pp. 137198-137208.

[19] van Bokkem, D., Hageman, R., Koning, G., Nguyen, L. and Zarin, N., 2019. Self-sovereign identity solutions: The necessity of blockchain technology. arXiv preprint arXiv:1904.12816.

[20] Sunyaev, A., 2020. Distributed ledger technology. In *Internet Computing*, pp. 265-299, Springer, Cham.

[21] Lux, Z.A., Thatmann, D., Zickau, S. and Beierle, F., 2020. Distributed-Ledger-based Authentication with Decentralized Identifiers and Verifiable Credentials. *arXiv preprint arXiv:2006.04754*.

[22] Patil, A.S., Belhekar, S.P., Burkul, R.S. and Sambare, M.V., 2019. Review Paper on–Smart Wallet.

[23] Laine, K., 2020. Homomorphic encryption. In *Responsible Genomic Data Sharing*, pp. 97-122, Academic Press.

[24] van Bokkem, D., Hageman, R., Koning, G., Nguyen, L. and Zarin, N., 2019. Self-sovereign identity solutions: The necessity of blockchain technology. *arXiv preprint arXiv:1904.12816*.

[25] Panait, A.E., Olimid, R.F. and Stefanescu, A., 2020, September. Analysis of uPort Open, an identity management blockchain-based solution. In *International Conference on Trust and Privacy in Digital Business*, pp. 3-13, Springer, Cham.

[26] Toyoda, K., Machi, K., Ohtake, Y. and Zhang, A.N., 2020. Function-level bottleneck analysis of private proof-of-authority ethereum blockchain. *IEEE Access*, *8*, pp. 141611-141621.

[27] Gilani, K., Bertin, E., Hatin, J. and Crespi, N., 2020, September. A survey on blockchain-based identity Management and decentralized privacy for personal data. In *BRAIN 2020: 2nd conference on Blockchain Research & Applications for Innovative Networks and Services*.

[28] S. Foundation, 2018, Sovrin: A protocol and token for self-sovereign identity and decentralized trust, Tech. rep., Sovrin Foundation.

[29] Ebrahimi, A., ShoCard, Inc., 2020, *BLOCKCHAIN ID CONNECT*. U.S. Patent Application 16/656,477.

[30] Dib, O., Brousmiche, K.L., Durand, A., Thea, E. and Hamida, E.B., 2018. Consortium blockchains: Overview, applications and challenges. *International Journal on Advances in Telecommunications*, *11*(1&2).

[31] Shah, M., Shaikh, M., Mishra, V. and Tuscano, G., 2020, June. Decentralized Cloud Storage Using Blockchain. In *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)*, *48184*, pp. 384-389, IEEE.

[32] Der, U., Jahnichen, S., And Surmeli, J., 2017, Self-sovereign identity − opportunities and challenges for the digital revolution, arXiv preprint arXiv:1712.01767.

[33] Ferdous, M.S., Chowdhury, F. and Alassafi, M.O., 2019. In search of self-sovereign identity leveraging blockchain technology. *IEEE Access*, *7*, pp. 103059-103079.

[34] Kuperberg, M., 2019, Blockchain-based identity management: A survey from the enterprise and ecosystem perspective. IEEE Transactions on Engineering Management, pp. 1–20.

[35] Zhu, X., And Badr, Y., 2018, Identity management systems for the internet of things: A survey towards blockchain solutions. Sensors, 18, 12, 4215.

[36] Berg, A., Berg, C., Davidson, S., and Potts, J., 2018, The institutional economics of identity. SSRN Electronic Journal, 05.

[37] El Haddouti, S. and El Kettani, M.D.E.C., 2019, Analysis of Identity Management Systems Using Blockchain Technology. In CommNet, pp. 1-7.

[38] Muhle, A., G Runer, A., Gayvoronskaya, T., and Meinel, C., 2018, A survey on essential components of a self-sovereign identity. Computer Science Review 30, pp. 80 – 86.

[39] Lesavre, Lo., 2019, A taxonomic approach to understanding emerging blockchain identity management systems, arXiv preprint arXiv:1908.00929.

[40] Van Bokkem, D., Hageman, R., Koning, G., Nguyen, L., And Zarin, N., 2019, Self-sovereign identity solutions: The necessity of blockchain technology, CoRR abs/1904.12816.