Self-Sovereign Identity: A Systematic Map and Review

FREDERICO SCHARDONG, Programa de Pós-Graduação em Ciência da Computação, Departamento de Informática e Estatística, Universidade Federal de Santa Catarina, Brazil and Instituto Federal do Rio Grande do Sul, Campus Rolante, Brazil

RICARDO CUSTÓDIO, Programa de Pós-Graduação em Ciência da Computação, Departamento de Informática e Estatística, Universidade Federal de Santa Catarina, Brazil

Self-Sovereign Identity is a user-centric identity model. In this model, the user maintains and controls their data. When requested by a service provider, user data is sent directly by the user, without the intermediation of third parties. Thus, in Self-Sovereign Identity, the participation of known identity providers for proof of identity is reduced, which increases user privacy. This identity model has attracted the attention of researchers and organizations around the world. All this interest increased the number of scientific articles published on the subject. The analysis of published materials showed that ideas and proposals are very diverse and dispersed. Although there are few systematic reviews, they lack methodological rigor and are limited to a small subset of published works. This study presents a rigorous systematic mapping and systematic literature review covering theoretical and practical advances in Self-Sovereign Identity. We identified and aggregated evidence from publications to answer four research questions, resulting in a classification scheme used to categorize and review publications. Open challenges are also discussed, providing recommendations for future work.

CCS Concepts: • **Security and privacy**;

Additional Key Words and Phrases: Self-Sovereign Identity, SSI, identity, identity management, identity and access management, privacy, systematic mapping, mapping, systematic literature review, literature review, review, survey

ACM Reference Format:
Frederico Schardong and Ricardo Custódio. 2021. Self-Sovereign Identity: A Systematic Map and Review. *ACM Comput. Surv.* 1, 1, Article 1 (August 2021), 35 pages. https://doi.org/XX.XXXX/XXXXXXX.XXXXXXX

1 INTRODUCTION

The ability to prove that a person is who she claims to be is fundamental to human interactions in society, whether in the physical world or on the Internet. The proof usually consists of presenting a credential, which allows the identification of a person. This credential is a set of attributes and is called an identity document or simply identity [19, 66].

Authors' addresses: Frederico Schardong, Programa de Pós-Graduação em Ciência da Computação, Departamento de Informática e Estatística, Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brazil, 88040-900 and Instituto Federal do Rio Grande do Sul, Campus Rolante, Rolante, Rio Grande do Sul, Brazil, 95690-000, frede.sch@gmail.com, frederico.schardong@rolante.ifrs.edu.br; Ricardo Custódio, ricardo.custodio@ufsc.br, Programa de Pós-Graduação em Ciência da Computação, Departamento de Informática e Estatística, Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brazil, 88040-900.

Manuscript submitted to ACM

We can classify identity documents into three different representation formats. The first format is the traditional physical document. This format usually consists of a paper document or plastic card, and the attributes that identify the person are printed on these materials. Paper and plastic cards are made with care to prevent them from being easily counterfeited. When a person wants to prove who she is, she presents her physical document. The identification is made by the relying party (RP) that wants to identify the person by reading and checking the veracity of the attributes. One of the crucial attributes of this type of document is the photo of the person's face. In this case, the identity document is known as a face badge.

The second format is a digital identity document. It can be perceived as the digital visual representation of the physical document and is often instantiated on mobile devices, *e.g.* [61]. Verification of integrity and authenticity is performed using cryptographic techniques such as digital signature. This signature and identity attributes are usually included as a QR code so that the RP can verify the integrity and veracity of this identity document.

The third format is the electronic identity document. This is the identity used in the virtual world, allowing people to authenticate themselves and consume electronic services on the Web. Unlike the digital document, which is a visual representation, the electronic document is designed from the ground up to be used electronically, eliminating visual inspection to verify its veracity and integrity. These processes are performed through multi-factor authentication [105] and cryptographic techniques such as digital signature and public-key cryptography [136]. For example, using a password that is only known to the identity holder and a key displayed in a time-based one-time password service [40, 150].

These three forms of identity need to be resistant to forgery, fraud, and personal information leakage. Therefore, the collection, storage, and processing of identity-related data must be done with great care, promoting the use of adequate data protection mechanisms. While each of the three forms of identity listed above is subject to fraud, the one that requires the most control is the electronic version. There are numerous reports of fraud involving the misuse of electronic identities [65, 96].

While there are examples of self-issued identity documents such as business cards and *curriculum vitae*, the vast majority of identities in use are issued by trusted third parties. For example, national-level identity documents such as driver's licenses, identity cards, and passports are commonly issued by the government [13] or private companies authorized for this purpose [51].

In today's digital world, big companies like Google and Facebook issue electronic identities. They created these identities to identify, authenticate, authorize and provide user attributes for their internal services. However, these identities have become a powerful tool to identify users not only to access the services provided by these companies, but also by various other service providers (SPs). Thus, these companies act as identity providers (IdP). Many companies have taken advantage of these IdPs to outsource the registration, identification, and authentication of their customers.

Using IdPs has many advantages and disadvantages. As an advantage, the user can have a single identity to authenticate to multiple SPs. A possible downside is that a single IdP holds data for a large number of users. Storing people's electronic identities in a few IdPs has been of great concern because these few data silos hold the data of many users [7]. These giant data silos have become attractive targets for hackers [56] as they contain high-value assets that can be misused [65] or even traded with institutions not authorized by users [96].

Although the vast majority of users naively trust IdPs, many users and companies are uncomfortable with the obligation to use and trust these entities. In this context, Self-Sovereign Identity (SSI) [7] has drawn attention, as it improves people's privacy, providing them with the means to store and manage their data. SSI has the potential to disrupt today's electronic identity ecosystem by not requiring people to actively use IdPs for their identification and authentication with RP.

Despite the provision of sovereignty over the digital presence, SSI presents new challenges that must be overcome for its wide adoption. The challenges are conceptual and pragmatic. The main conceptual challenge is understanding what a self-sovereign identity is and what constitutes a self-sovereign system. The pragmatic challenges include, but are not limited to, how to coexist and migrate the identity stored and managed in existing IdPs to the new model, how to trust data from other self-sovereign identities, and how to help the user handle the tasks of management, backup, and recovery of private data.

The advantages of this new identity paradigm over traditional models has attracted the attention of researchers and professionals in recent years, which has led to a growing number of publications on the subject. There are also some initiatives to review and condense the knowledge produced so far. However, the existing reviews do not cover all the aspects currently considered important on the subject. For example, they fail to: (i) capture publications that contribute to the conceptual debate about what the term self-sovereign identity means; and (ii) include efforts that introduce new problems and solutions, but that are not implementations that cover the entire scope of the subject, such as Sovrin [140] and uPort [146]).

This article presents a comprehensive systematic mapping and review of scientific and non-scientific works that contribute to the debate about what SSI means and works that present a practical challenge concerning SSI and offer a solution for it. We followed a systematic method of searching and selecting publications, which was guided by four research questions. As our work is systematic, it may be reproduced and updated in the future to cover new work. Our results consist of an analysis of publication frequency, venues, co-references and co-authorships, and three maps that map the reviewed literature, providing the reader with an overview of the state of the art of SSI literature. Finally, open challenges and recommendations are discussed, which are useful for researchers and practitioners working with SSI.

The rest of this article is organized as follows. Section 2 provides background on electronic identity and a detailed description of SSI. The existing secondary works that review the SSI literature are provided in Section 3. The methodology adopted in this study is described in Section 4, and the results are presented in Section 5. Finally, we discuss the open challenges and shortcomings in Section 6 and final remarks in Section 7.

2 PRELIMINARIES

In this section, we present the background necessary to accompany this study. We start by explaining electronic identities, how their use has evolved, and then we detail SSI.

2.1 Electronic Identity

In the physical world, trust in relations between different entities involves identifying the communicating parts. Proof of identity is carried out through authentication factors previously agreed between the parties or with the assistance of trusted third parties. Typically, physical devices are used as authentication factors. For example, it is not uncommon for people to be identified through a visual inspection of their identity document, followed by a facial badge verification. Likewise, in the electronic world, communicating parties need to have a certain level of assurance of the other parties' identity. This assurance is done through data communication networks such as the Internet, using electronic identities.

Like a physical identity, an electronic identity is usually defined as a set of attributes that help describe or qualify an entity [66]. Some authors prefer to restrict this definition to specific contexts to improve its accuracy [35, 44, 98]. Consequently, electronic identities are not just copies of physical identities, such as a passport or driver's license. They are created, used, and destroyed according to the user's desire, often containing only the minimum attributes necessary to accomplish what is needed in that context. For example, a seller may have an electronic identity on eBay [34] without revealing their name, age, or country of residence. The only information that concerns others is whether this seller has a history of positive transactions [110].

All identities, whether physical or electronic, are subject to the existence of mechanisms to verify ownership. That is, they need mechanisms to carry out the appropriate *identification* and *authentication* of users [72].

The identification process consists of an electronic identity holder showing a unique attribute in a given context, *i.e.*, an identifier used to distinguish it from all other electronic identities in that context [46]. The classic example is providing an email address when subscribing to a subscription service. The next step is to authenticate the identified entity by verifying a security proof, traditionally performed through a secret password or digital signature, thus ensuring that the holder of the electronic identity is its owner *de facto*. In the example of the subscription service given above, providing a code or clicking a link received by email provided proves that the email address belongs to the holder.

Performing identification and authentication is crucial in our digital society, allowing services to be made available electronically to citizens. Therefore, this identification and authentication activity should be done by a specialized service trusted by all involved entities. Such services are provided by systems that govern electronic identity, so-called identity and access management (IAM) systems.

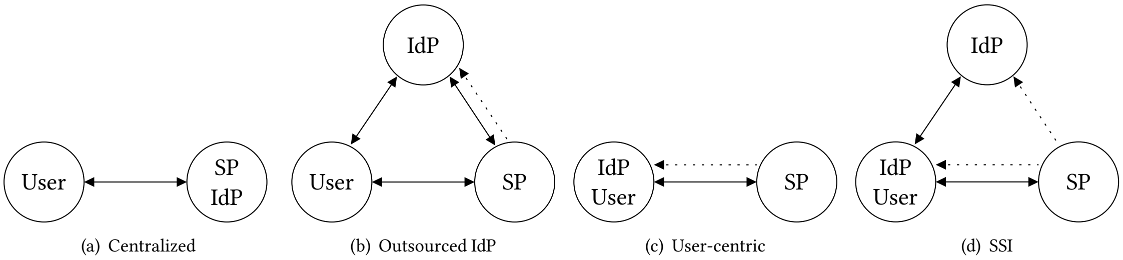
2.2 The evolution of IAM models

In the early days of the web, SPs had to implement IAM solutions to identify and authenticate clients to offer personalized products and services. Thus, such services are called *centralized authorities*. This model, however, presented a series of difficulties for users concerning usability. Most users ended up using similar low-entropy passwords on different systems, making room for numerous vulnerabilities. This model has spurred many efforts to educate users about the dangers of having simple passwords and reusing them across different services [114, 139].

The next logical evolution replaced the centralized model with third-party IAM solutions, IdPs. With this new paradigm, users need to be registered in just a few IdPs to access the plethora of services available on the web. On the other hand, SPs must be registered with the desired IdPs or federations of IdPs to work with identified and authenticated users of the IdPs. The interactions between IdP, SP, and the end-user were standardized through token exchange protocols like SAML [59], OAuth [57] and OpenID [108]. Although this identity model has dramatically reduced the problematic management of various identifiers and passwords for users, it ended up creating few and large silos of valuable private information.

The next evolution step was the user-centric model [69]. It was built on the idea that users could use personal authentication devices (PADs), such as smartphones and smartcards, to store authentication credentials from SPs and decide which ones to present, thus eliminating the need for third-party IdPs. However, as pointed in [7], this model has not gained momentum and is currently understood as the IdP model with greater user control. The current understanding of this model, according to [7], is that the user is aware and must authorize or deny her IdP to share specific personal attributes requested from an SP. Therefore, the current user-centric identity model presents the same issues as the previous model.

The Figures 1(a), 1(b) and 1(c) depict the identity models presented above, providing a high-level overview of interactions between user, IdP, and SP. The creation of specialized IAM services, *i.e.*, third-party IdPs, resulted in the emergence of oligopolies of electronic identities [62]. Long-term users of IdPs are effectively imprisoned by them since there is no portability between IdPs. These companies promote their own rules, which, if violated, can remove a user from their platforms. This can be disastrous for individuals who spent years building trusting relationships with SPs, and if banned, they would lose their transaction history and become completely unknown. This problem is of particular importance in the case of IdPs that are also social media platforms, such as Facebook, LinkedIn, and Twitter, where violations of rules on the social network are often questionable [47].



(a) Centralized (b) Outsourced IdP (c) User-centric (d) SSI

Fig. 1. The IAM models. Constant lines represent interactions, and dashed lines mean trust.

2.3 Self-Sovereign Identity

In the early days of the web, the conception of the client-server model shaped the idea that in the digital world, people are users of online systems rather than human beings, *i.e.*, entities that need identification, authentication, and authorization to access and perform tasks online [121]. This digital model assumes administrative precedence [90] because it was built on the foundation that servers (companies, online businesses) are more important than clients (individuals) and, therefore, dictate the rights of clients. This web fabric holds to this day and is exacerbated by the need to the creation of legislation, such as the European Union's General Data Protection Regulation (GDPR) [41], and the California Consumer Privacy Act (CCPA) [27], to specify the rights of individuals and their digital data in a society increasingly dependent on digital interactions.

The core concept of SSI is that individuals are sovereign over their digital selves and, therefore, have control of their data. This idea fundamentally differentiates SSI from previous identity models, in which individuals were seen as users. In this new model, sovereign individuals store and manage their data, thereby controlling who their private data are shared with and to what extent.

Although philosophers such as John Locke and Stuart Mill have written about the sovereignty of individuals in past centuries [97, 147], the first commonly accepted [7, 125, 159] connection between the concepts of sovereignty and digital identity was made by Leffreto [89], where he stated that "Individuals have an established right to an "identity"" [89]. Thereafter, the meaning of sovereign identity was debated [91, 122–124] and technology standards were proposed [109, 132]. Significant momentum was obtained, especially in academia [102, 143], after Christopher Allen laid out what he proposes to be the ten principles of SSI [7], which are detailed next.

First, individuals must have an **existence** independent of their digital selves, *i.e.*, they cannot exist only virtually. A (self-sovereign) identity works by sharing the desired (digital) aspects of the individual. Second, people must **control** their identities, owning and managing their attributes, which does not prohibit them from making *claims* about other people. Third, people must have **access** to their data and claims by storing them or being readily available if they are outsourced. Fourth, all systems must be **transparent** and the underlying algorithms must be free and open-source, thus allowing detailed examination by anyone. Fifth, identities must **persist** forever, or as long as individuals wish. Sixth and seventh, identities and their claims must be **portable** across different systems and technologies, which requires **interoperability** between standards and implementations. Eighth and ninth, people need to **consent** to the use and sharing of their data, while disclosure of data must be **minimized** to the absolute minimum. For instance, to find out if a person can buy an alcoholic beverage, it is unnecessary to share the date of birth. Tenth, at the end of the day, individuals must have their rights **protected**, which means that systems must be designed to avoid censorship and must preserve people's rights, even to the detriment of systems.

In SSI, any assertion regarding a subject is called a **claim**. A set of one or more claims made by an entity despite a subject is a **credential**. It could be, for instance, a government-issued driving license containing the date of birth, name, and address of a person. Suppose the credential is associated with a revocation list or other revocation method and has cryptographic material assuring its integrity and identification and non-repudiation of its issuer. In that case, it is referred to as a **verifiable credential** (VC) [132]. Furthermore, a tamper-proof claim derived from a verifiable credential is called a **verifiable claim**. Although we will use these terms interchangeably in the remainder of this paper, we will be referring to tamper-proof claims and tamper-proof credentials unless specified otherwise.

Similar to the physical world, where entities issue physical credentials to holders in paper or plastic cards, in SSI, entities issue VCs to holders. However, different from physical and digital identities, these electronic identity documents allow individuals to choose what attributes (claims) they want to share, which is impossible with physical and digital credentials. They require the holder to present the entire identity document, thus revealing all its attributes.

Suppose that you are asked to prove to have reached the age of majority. With a physical document, it is evident that showing the paper or plastic card will reveal the birthdate and all other attributes to the RP. The same is true for digital identity documents, which are commonly implemented using X.509 attribute certificates [145]. With traditional X.509 certificates, the whole certificate has to be shared with the RP to verify the document's integrity. However, in the context of SSI, you would construct a **verifiable presentation** (VP) stating that: (i) a credential was issued to you by a trusted party; (ii) this credential has your birthdate in it; (iii) your birthdate was more than 18 years ago; and (iv) this credential have not been revoked by the government body. Hence, whoever receives this VP does not learn your name, birthdate, and any other information in the credential, only that you have reached the age of majority.

The recipient of a VP (*i.e.* the RP) checks: (i) who signed the credential that supports this VP; (ii) if the VP is correctly constructed (*i.e.* it contains the required information and is not corrupted nor counterfeited); and (iii) whether the credential supporting this VP is valid (*i.e.* whether the credential was revoked or not). It is important to note that upon verifying who issued the credential on step (i), the RP is free to decide if it trusts the issuer or not. Moreover, step (iii) does not require the RP to inquire the IdP in any particular manner. Revocation registries are publicly available, and the verification is done anonymously [24, 25], *i.e.*, not revealing a unique identifier of the credential.

Although individuals are autonomous to issue credentials for themselves in SSI, others are free not to trust them. For instance, a bank will hardly accept a VP of a self-issued credential containing a person's name and birthdate. This is true both in the physical world and in SSI. Figure 1(d) provides a high-level overview of SSI, where the user (*i.e.*, the holder) can interact with the SP using self-issued credentials or third-party issued credentials. In both cases, the SP must trust the issuer.

Despite the terminology adopted by the SSI literature regarding VP, this concept precedes SSI for many years. Research has been carried out for more than a decade before SSI on how to share part of a credential, as well predicates over one or more attributes of a credential, without losing integrity and authenticity [24, 28]. The primary technique underlying VP is zero-knowledge proof (ZKP) [117]. Generally speaking, a ZKP enables a prover to convince a verifier that she knows a value without revealing the value [119]. Using ZKP schemes with credentials, a credential holder can prove the validity and content of one or more credentials without revealing the whole credentials [81]. The same is true for the status of a VC. It is possible to prove that a VC has not been revoked without revealing to the RP an identification of the credential and ensuring that the issuer does not know that a query for a specific credential was made [25].

In Figure 2 we present a generalized three-actor overview of the end-to-end process of issuing a VC and emitting a VP. In this example, three people own and control their electronic identities, each suitable for a different context. Each electronic identity has a database of issued and received credentials, and a revocation registry for issued credentials that have been revoked. One of Alice's electronic identities issues a credential to one of Bob's electronic identities, for instance, a statement that he is a trusted seller of good wines. Bob then sends a VP of having such credential to Carl's electronic identity used to search and buy wine. Carl, in turn, trusts the issuer of the credential in which that VP was derived (Alice, an internationally recognized winemaker) and proceeds to negotiate with Bob. It should be noted that, in real life, most people will not host revocation registries as they will not issue credentials, which is also the reality with physical and digital identification documents.

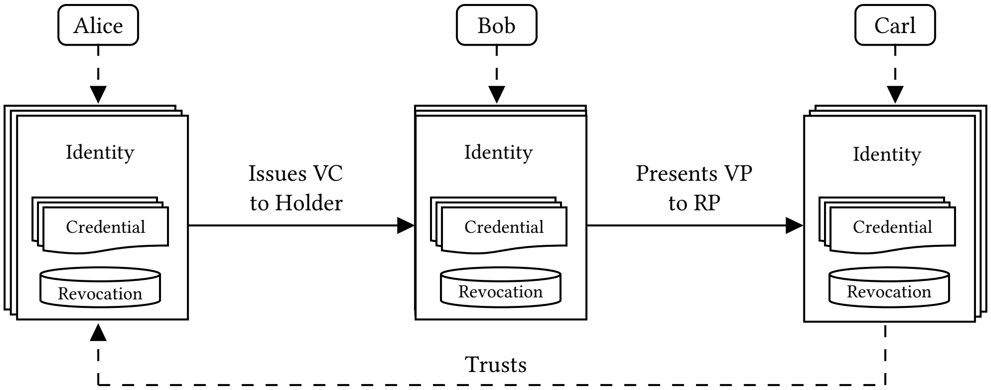


Fig. 2. The actors, their electronic identities, and the interactions to issue a credential and present a VP.

After introducing electronic identities and the evolution of IAMs, we presented SSI to the reader. Next, we present other surveys of SSI and their shortcomings in Section 3, followed by the methodology used in this systematic mapping and systematic review in Section 4.

3 RELATED WORK

Blockchain technology introduced the idea of distributed ledgers, where consensus among peers defines the state of the immutable ledger, rather than a central entity with authoritarian control [138]. These concepts facilitate the implementation of SSI by providing a trusted online storage of digital identities, credentials, and revocation registries [78]. However, while blockchain assists the development of SSI solutions, it is not a requirement [6, 8, 84, 148]. Nonetheless, existing reviews claim that SSI necessarily needs blockchain to be implemented [32, 48, 71, 78, 83, 86, 102, 162]. With regards to the review methodology of these works, two surveys provided details about their search strategy [78, 86] while six did not [32, 48, 71, 83, 102, 162]. Next, we present the existing secondary studies in the field of SSI.

Kuperberg [78] conducted a systematic review where 43 blockchain-based SSI market offerings were assessed using 75 criteria, covering legislation compliance, market availability, costs, and more. He argued that no reviewed application meets all criteria and no SSI solution has: (i) the maturity of conventional IAM offerings; (ii) a production-level integration standard (such as OAuth [57] and SAML [59]); and (iii) OS-level integration.

Liu *et al.* [86] systematically reviewed 36 research efforts and patents that introduce SSI applications. They examined these works based on authentication, privacy, and trust aspects. They argued that, despite blockchain-related innovations, there are still issues and implications remaining, namely: (i) users may lose their blockchain-based identities (wallets) and need to (ii) change their identities, which is trivial in traditional IAM but might be challenging in distributed ledgers; and (iii) the cost of integrating existing systems to the new paradigm.

Focusing on the internet of things (IoT), Zhu and Badr [162] surveyed works that use distributed ledgers to provide SSI in the context of IoT devices. They shared Liu *et al.*'s [86] focus on authentication, privacy, and trust issues and added a fourth: performance. They argued that the trustless environments of IoT devices naturally demand SSI solutions. Still, blockchain technology must be carefully explored as storing and maintaining public blockchains in IoT devices is resource-prohibitive. Therefore, having small groups of private blockchains could be an alternative. A possible solution reported from the literature [161] is that IoT devices could inherit the peer-to-peer trust between their owner entities (human, business, government).

Three surveys that do not specify a search method have produced similar outputs despite the comparison of the underlying infrastructure of blockchain-based SSI offerings [48, 71, 83]. They all highlighted what blockchain framework the surveyed work use, the type of the blockchain network (*i.e.* if they are private, permissioned, permissionless, or other). Lim *et al.* [83] reviewed 15 for-profit and non-profit company-made, government-related, and open-source applications, they highlighted that SSI presents the ideal solution for proper user-centric, secure and cost-effective IAM. Kaneriya and Patel [71] reviewed 6 SSI systems, enumerating future enhancements that, according to the authors, each system should focus on. Lastly, Gilani *et al.* [48] reviewed 8 SSI offerings, detailing which support selective disclosure of personal information, how cryptographic keys are managed, and blockchain-specific details such as whether credentials are stored either on or off ledger and the usage of smart contracts. Smart contracts are a piece of software that runs on the ledger automatically and transparently for everyone to verify [160].

The authors of [32] described 10 SSI systems that use blockchain without specifying how these were selected. Nonetheless, they conducted an analysis of these works with regards to their attendance to the ten principles of SSI [7], individually detailing which principle each reviewed paper satisfies.

Differently from the previous surveys, Mühle *et al.* [102] reviewed what the authors call "four basic components of SSI": (i) identification; (ii) authentication; (iii) verifiable claims; and (iv) attribute storage. They discussed how different research work and market offerings try to provide solutions for each of the four components.

Secondary work already carried out shows that there is a growing number of studies in this area. However, there is a need for more systematic identification of SSI studies, regardless of technological implementations. Previous surveys investigate practical and technical aspects of SSI systems. Nevertheless, they do not evaluate conceptual discussions about SSI nor works that present and try to solve pragmatic challenges that are not entire SSI systems. In this work, on the other hand, we are interested in systematically discovering and investigating research materials that: (i) extend or refute Allen's ten principles of self-sovereign identity [7]; (ii) introduce definitions or formalization to SSI; or (iii) present and solve practical problems in the SSI ecosystem. The key differences between previous surveys and ours are summarized in Table 1.

Table 1. Comparison with other surveys in the literature.

	Systematic Review	Systematic Mapping	Other than Blockchain	Conceptual or Pragmatic	Covered Works
Kuperberg [78]	Yes	No	No	Pragmatic	43
Liu *et al.* [86]	Yes	No	No	Pragmatic	36
Zhu and Badr [162]	No	No	No	Pragmatic	15
Lim *et al.* [83]	No	No	No	Pragmatic	15
Kaneriya and Patel [71]	No	No	No	Pragmatic	6
Gilani *et al.* [48]	No	No	No	Pragmatic	8
Dib and Toumi [32]	No	No	No	Pragmatic	10
Mühle *et al.* [102]	No	No	No	Pragmatic	9
This work	**Yes**	**Yes**	**Yes**	**Both**	**57**

4 METHODOLOGY

As the primary research efforts surrounding a given topic evolve, secondary studies are needed to keep track of advances and developments. In computer science, two types of secondary studies became popular in recent years [107]: (i) systematic mapping [106]; and (ii) systematic literature review [75]. Despite both being systematic and thus follow a protocol/methodology to identify and interpret relevant research work, the former is designed to produce a broad overview and to discover research trends, while the latter is targeted to aggregate evidence to summarize and answer narrower research questions. In this study, we attack both challenges.

We followed the methodology introduced by Petersen *et al.* [107], which provide detailed guidelines after having systematically surveyed mapping studies. These guidelines demand for: (i) the definition of objectives and research questions (section 4.1); (ii) a strategy to identify relevant studies (section 4.2); (iii) objective inclusion and exclusion criteria to ensure only relevant material is reviewed (section 4.3); (iv) an extraction process to impartially retrieve evidences from papers regarding the research questions (section 4.4); (v) to present the reader how selected materials will be classified (section 4.5); and (vi) the discussion of possible threats to the validity of the study (section 4.6).

4.1 Research Questions

The objective of this systematic study is fourfold: (i) to identify and map SSI-related publications; (ii) to investigate what are the conceptual advances made to the informal definition of SSI [7]; (iii) to explore mathematical formalism of SSI; and (iv) to examine practical problems and solutions related to SSI. These objectives lead to the following research questions (RQs):

- RQ-1: When, where, and by whom were SSI studies published?
- RQ-2: What conceptual ideas have been introduced or refuted?
- RQ-3: What properties or formal definitions have been specified?
- RQ-4: What practical problems have been introduced and solved?

4.2 Search Strategy

The first step in our search was to specify a search string relevant to the research questions mentioned above. Rather than creating a potentially restrictive search query following PICOC [75] or any other method to frame search queries, we searched for “self-sovereign identity” and variations in the title, author keywords, and abstract. Our search string is general on purpose. We wanted to include as many relevant articles as possible. Furthermore, we applied no restrictions on the year of publication, the number of pages, or conference/journal. Our complete query string is as follows.

self-sovereign identity **OR** self sovereign identity **OR** self-sovereignty **OR** self sovereignty

On June 25, 2021, this search string was queried on ACM Digital Library [11], IEEE Xplore [64], ScienceDirect [37] and Springer Link [135], which are databases hosting popular computer science conferences and journals. To expand the reach of our database search, we also searched the Scopus Preview [39], Web of Science [29] and Google Scholar [49]. Additionally, we used Google Patents [50] and applied the search string on the title and abstract of patents on the same day, finding 11 results. Table 2 shows the number of search results returned from the queries.

Table 2. Number of studies.

Tool	**Total**
ACM Digital Library	8
IEEE Xplore Digital Library	61
ScienceDirect	5
Springer Link	27
Scopus	168
Web of Science	77
Google Scholar	114
Database Search	**460**
Google Patents	11
Patent Search	**11**

We then submitted the articles and patents to the study selection process described below, producing an initial set of works to be reviewed. This set of results served as input to a snowballing process, which helps to ensure that any relevant publication was included in this review.

4.3 Study Selection

The study selection process has three phases. In the **first phase**, both duplicate results and articles republished in extended formats are removed. The software Mendeley [38] was used to discover and remove the duplicates.

A first screening of the search results revealed that some papers do not belong to the field of computer science or are outside the scope of our review. We then created two inclusion criteria and one exclusion criterion to narrow our search. Table 3 describes these criteria. In short, the exclusion criterion filters out research work outside of computer science, while the inclusion criteria select papers that make contributions to SSI according to our research questions. Articles need to satisfy at least one inclusion criteria.

It is important to note that, despite arguably incorporating *practical progress* (w.r.t. IC-2), we are not mapping and reviewing standalone SSI solutions (such as Sovrin [140] and uPort [146]). Although multiple surveys [78, 83, 86, 162] focus on these works, they aim to tackle an issue not yet precisely defined. Therefore, when it comes to *practical progress*, we focus on works that present specific pragmatic concerns related to any facet of the SSI ecosystem and offer a solution. For instance, consider an article that points out the challenge of recovering lost keys in SSI and introduce a novel solution for this problem. This work would be in accordance with IC-2. However, suppose a research paper

Table 3. Inclusion and exclusion criteria.

Inclusion Criteria	
IC-1	The paper includes a novel conceptual contribution to SSI.
IC-2	The research work makes practical progress towards SSI.
Exclusion Criterion	
EC-1	The research work is not in the area of computer science.

reporting some implementation of SSI for IoT. While it might be a relevant contribution to the IoT literature, if it does not present a problem concerning SSI in general and a solution for such problem, then this work does not satisfy IC-2.

In the **second phase**, EC-1 is applied to the title, author keywords, and abstract, hence removing articles that are not in the area of computer science. Next, in the **third phase**, the remaining studies are obtained and read in full, observing their compliance with IC-1, IC-2, or both. Then, articles that do not comply with IC-1 nor IC-2 are also removed.

Our three-phase study selection process was executed five times, as presented in Figure 3. The first execution was regarding the outputs of the database search, and the second with the results of the patent search. The combined result of these two executions was a set of 42 works used as the basis for both forward and backward snowballing [157]. In short, backward snowballing consists of reviewing all citations in a document, while forward snowballing finds other works that referenced it. The snowballing was repeated until no new work was found that satisfied our selection process, which required three runs. The remaining 57 articles constitute our result set. We should point out that each paper was independently assessed by two researchers in every stage of the selection process, and a conflict resolution meeting was organized. We point the interested reader elsewhere [118] for a list of all papers and our evaluation regarding their inclusion or exclusion for all five runs of the study selection process.

4.4 Data Extraction

We adapted the Petersen *et al.*'s [107] template to extract data from primary studies. It consists of a series of: (i) data item; (ii) a description, and; (iii) the RQ that it responds to, as shown in Table 4. The *General* items were obtained from the articles or their metadata available online, except for the Study ID, which was generated manually. After reading a pilot set of articles, two *Conceptual* and two *Practical* data items were created to collect evidence and answer RQs. The articles and patents returned from the database search and snowballing, together with the evidence collected for completing the data extraction form, can be found elsewhere [118].

4.5 Classification

Petersen *et al.* [107] advocate reusing classification schemes from previous studies, claiming that it allows a direct comparison between two mappings. To the best of the authors' knowledge, this is the first systematic mapping of SSI. Therefore, we present a new classification strategy. We created this classification following Petersen *et al.*'s [106] *keywording* methodology, which has three steps: (i) the researcher reads the abstracts (introduction and conclusion too, if the abstract is of low quality), extracting keywords and concepts that indicate the contribution of the article and the context of the research; (ii) the set of keywords is combined to constitute a high-level understanding of the research contribution; and (iii) the final set of keywords is clustered to form the map categories. The last step is the

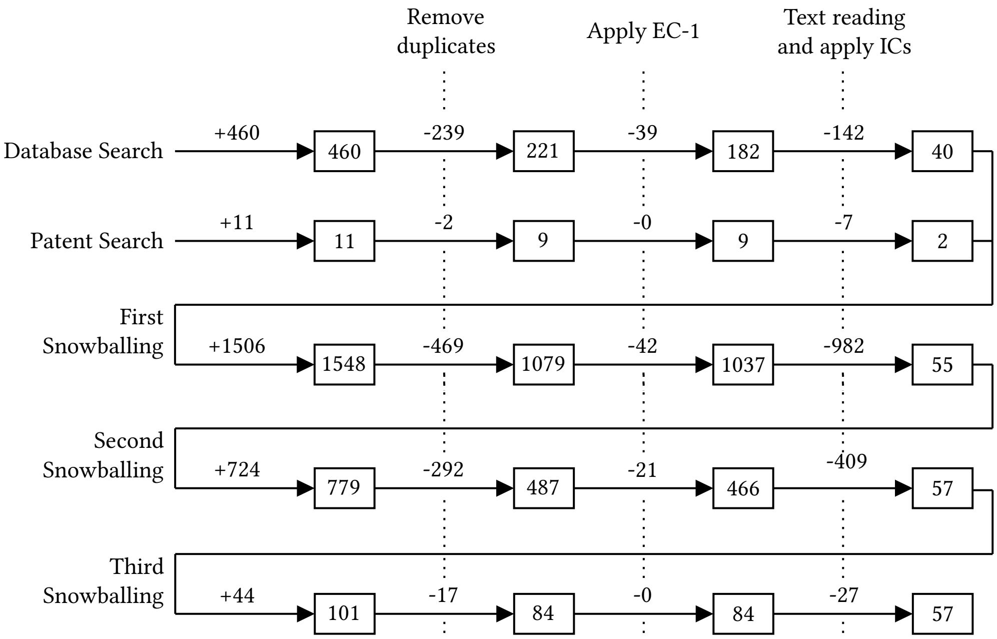


Fig. 3. Number of articles in each stage of our search strategy and study selection.

Table 4. Data extraction form adapted from [107].

Data Item	Description	RQ
General		
Study ID	Unique integer identifier per article	
Article Title	Name of the article	
Year	Year of publication	RQ-1
Article Authors	Name of the authors	RQ-1
Venue	Publication venue	RQ-1
Conceptual		
Add Concept	What concept/idea is introduced	RQ-2
Refute Concept	What concept/idea is refuted	RQ-2
Formalism		
Formal Model	How is SSI formally specified	RQ-3
Practical		
Novel Problem	What practical problem is presented	RQ-4
Proposed Solution	How is the practical problem solved	RQ-4

result of the continuous act of creating, updating, and merging categories, together with the classification of articles to the new categories.

4.6 Threats to Validity

The following threats to validity are of great importance and need to be pointed out [107]: (i) descriptive; (ii) study identification; and (iii) data extraction and classification.

To reduce the threat of inaccurately collecting observations from research papers (descriptive validity), we have developed and used the data collection form described above to collect relevant evidence. The first author applied the data collection form, and then these results were evaluated by the second author.

Next, to mitigate the possibility of missing relevant work (study identification validity), we did not restrict our database search by year of publication or venue. Besides, the database search was supplemented with backward and forward snowballing.

Researcher bias and human error during data extraction and classification cannot be completely eliminated, as these processes are matters of human judgment. The second author evaluated the data extraction and classification performed by the first author to reduce these threats.

Furthermore, it is essential to mention that identity management has been studied for decades. Therefore, numerous research efforts can arguably contribute to the many facets of SSI (*e.g.*, trust, key management, certificate) despite being conducted before the term self-sovereign identity was conceived. Ultimately, deciding which of these works may contain a valuable contribution to SSI is limited by the researcher's interpretation. Therefore, to suppress this interpretation bias, especially in the snowballing process, we mapped and surveyed only works that explicitly mention the term self-sovereign identity (or a synonym like self-sovereignty).

5 RESULTS

In this section, we present in five parts the outcomes of our systematic mapping and review. First, we describe the results of the keywording strategy to devise the classification, and these serve as a basis for reporting the data collected concerning our research questions. Then, based on our categorization, the four research questions are addressed.

5.1 Classification

The keywording methodology [106] resulted in two Venn diagrams, shown in Figure 4. Unlike taxonomies, in which facets are mutually exclusive, Venn diagrams allow classifying publications with different levels of granularity. This less rigid form of mapping presents an ideal starting point for categorizing research in SSI. Future secondary work can build upon our model to produce taxonomies or other types of classifications.

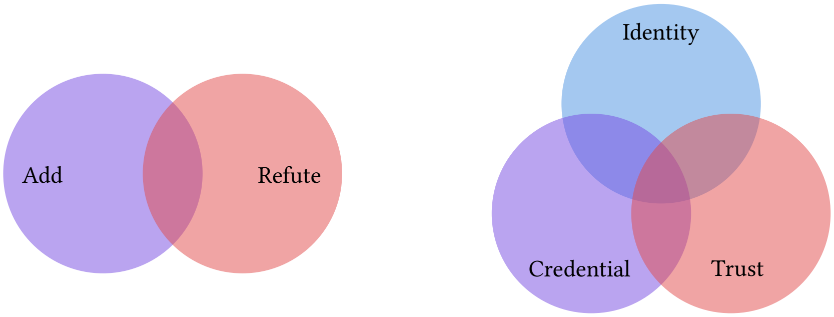


Fig. 4. The two Venn diagrams that form the basis of our maps.

The diagram on the left side of Figure 4 is used to map the research efforts that in our data extraction process have filled in the data items *Add Concept* or *Refute Concept* and thus help answer RQ-2. On the other hand, the diagram on the right is used twice. First, it is used to map publications that have completed the data item *Formal Model*, in relation to RQ-3. Then, it is used to map publications with pragmatic contributions, *i.e.*, that filled the data items *Novel Problem* and *Proposed Solutions*, therefore in relation to RQ-4.

The left diagram categories (**Add** and **Refute**) and the right diagram categories (**Identity**, **Credential** and **Trust**) are the starting points for our classification. In the following sections, these diagrams are augmented, and new intersections are drawn, which increases the depth and level of detail, thus creating the maps that answer our second, third, and fourth research questions.

While the distinction between adding and refuting concepts is clear, the boundaries between the categories in the diagram on the right are not. For a work to be placed entirely in the identity category, it must not involve issuing, revoking, storing credentials, or presenting claims, nor challenges related to establishing trust between identities. Articles on these two subjects are classified in the credential and trust categories, respectively. However, if a researched paper, for example, primarily addresses key management and claim presentation, it would be classified in the intersection of the identity and credential categories.

5.2 When, where, and by whom were SSI studies published?

To answer RQ-1, we aggregate the *General* data items collected using our data extraction form. The results are discussed next.

5.2.1 Frequency of publication. Regarding publication frequency, we aggregate publications by year in Table 5. Although it provides a rough overview, the growing academic interest in SSI is evident. In addition, we see finer details by applying the classification described above to annual publication frequency. Figure 5 shows the number of publications according to our classification.

Table 5. Frequency of publication per year.

Year	Total	Studies
2016	3	[7, 9, 120]
2017	5	[1, 33, 132, 149, 151]
2018	5	[52, 84, 102, 116, 137]
2019	13	[16, 36, 45, 53, 54, 81, 92, 109, 117, 129, 130, 143, 154]
2020	18	[2, 3, 14, 20, 58, 63, 67, 74, 85, 87, 93, 104, 112, 115, 131, 144, 156, 158]
2021	13	[4, 5, 55, 68, 73, 77, 79, 80, 82, 88, 103, 127, 128]

Three publications were made in 2016, including the introduction of the term SSI and the ten principles by Allen [7]. Furthermore, both works published in 2016 and 2017 consist of conceptual writings that extend the discussion introduced by Allen, proposing new principles/requirements for SSI [1, 7, 9, 33, 120, 149] together with the coherent refutation of some [120]. Starting in 2018, research efforts began to introduce new pragmatic problems and solutions to the SSI ecosystem, along with mathematical formalism that, although fewer in number, help mature SSI towards a well-defined area of research.

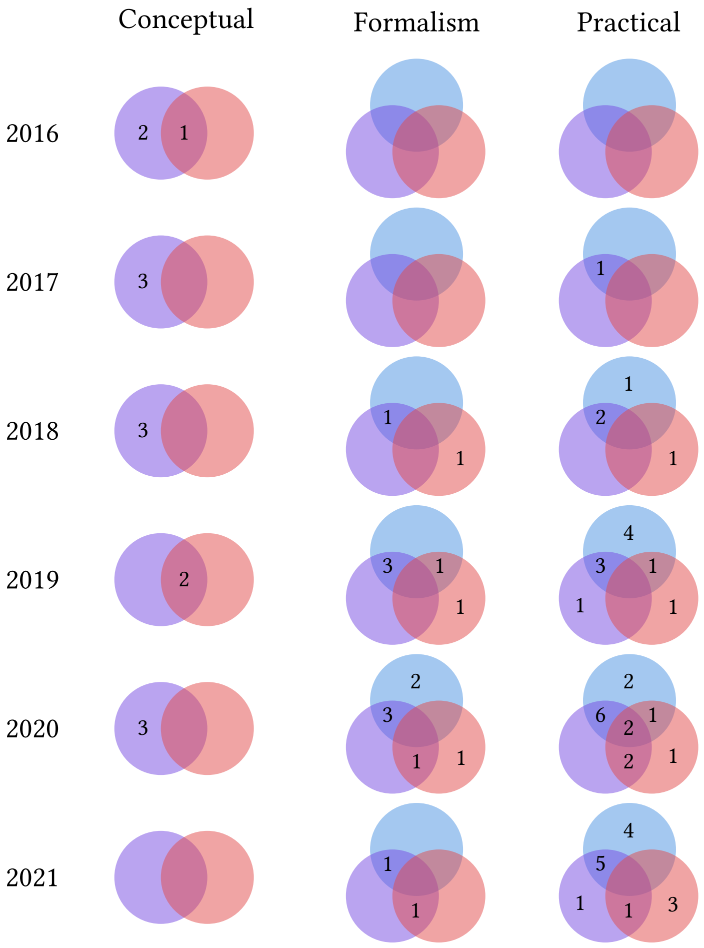


Fig. 5. Number of publications over the years divided by categories.

5.2.2 Publishing Venues. On the subject of publication venues, most publications (thirty articles) took place in congresses, symposia, forums, or workshops as shown in Table 6, which includes them under the term conference. Having thirty articles in conferences and four master's thesis shows that SSI has been gaining momentum as a research field. However, it is in its infancy, as there is only one doctoral thesis and only nine articles in journals.

The conferences selected by authors to publish their works vary widely. Table 7 shows in which conferences, symposia, forums, and workshops the mapped articles were published. Although there are thirty articles mapped in this type of venue, only six conferences received two publications. The other eighteen papers are spread across eighteen different conventions. Moreover, the colloquia held by the IEEE are the most popular choice, which received thirteen papers spread across ten conferences. The same trend is true for essays in scientific journals, as shown in Table 8. Five of the nine studies were published in IEEE journals.

5.2.3 Authors. Having the authors' names collected through our data extraction form, Figure 6 shows a co-authorship network [113] created using Gephi [15]. Each vertex represents an author and edges map published works between

Table 6. Types of publishing venues over the years.

Venue Type	Total	2016	2017	2018	2019	2020	2021
Blog Post	2	[7, 120]					
Tech Report	2		[1, 151]				
Standard	3		[132]		[109, 154]		
Web Archive	4				[81, 117] [129]	[14]	
Patent	2						[55, 103]
Conference	30	[9]		[52, 137] [84, 116]	[16, 53, 54] [92, 130]	[2, 67, 112] [74, 104, 156] [85, 93, 144] [3, 20, 63]	[5, 77, 128] [73, 79, 82] [68, 127]
Journal	9			[102]	[45, 143]	[58, 87, 158]	[4, 80, 88]
Master Thesis	4		[33, 149]		[36]	[131]	
PhD Thesis	1					[115]	

Table 7. Studies published in conferences, symposia or forums.

Venue Name	Total	Studies
Conference on Blockchain Research & Applications for Innovative Networks and Services	2	[85, 93]
IEEE International Congress on Cybermatics	2	[52, 137]
IEEE International Conference on Blockchain and Cryptocurrency	2	[68, 79]
IEEE International Conference on Trust, Security and Privacy in Computing and Communications	2	[3, 116]
International Conference on Information Networking	2	[73, 82]
Open Identity Summit	2	[5, 77]
IEEE International Conference on Internet of Things: Systems, Management and Security	1	[92]
IEEE International Conference on Mobile Cloud Computing, Services, and Engineering	1	[104]
IEEE International Symposium on Network Computing and Applications	1	[53]
IEEE International Symposium on Dependable, Autonomic and Secure Computing	1	[130]
IEEE Symposium Series on Computational Intelligence	1	[54]
IEEE International Conference on Pervasive Computing and Communications Workshops	1	[127]
IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops	1	[16]
IFIP International Conference on Information Security Theory and Practice	1	[67]
IFIP International Summer School on Privacy and Identity Management	1	[112]
IFIP International Conference on New Technologies, Mobility and Security	1	[128]
International Conference on Information and Communications Security	1	[2]
International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing	1	[74]
Symposium on Cryptography and Information Security	1	[156]
Annual Conference of the South African Institute of Computer Scientists and Information Technologists	1	[84]
International Teletraffic Congress	1	[63]
International Symposium on Networks, Computers and Communications	1	[20]
Annual Privacy Forum	1	[144]
Rebooting the Web-of-Trust	1	[9]

them in this weighted undirected graph. For ease of reading, we display edge weights in different line widths. Vertexes also change in diameter, representing the number of publications each author has. The vast majority of edges of this

Table 8. Studies published in journals.

Journal Name	Total	Studies
Frontiers in Blockchain	2	[4, 88]
IEEE Access	1	[45]
IEEE Software	1	[87]
IEEE Security and Privacy	1	[143]
IEEE Transactions on Vehicular Technology	1	[158]
IEEE Transactions on Computational Social Systems	1	[80]
Elsevier Computer Science Review	1	[102]
MDPI Electronics	1	[58]

network graph are thin, showing that most authors have a single publication. In addition, this disconnected graph also indicates that most of the work was done alone or in isolated groups.

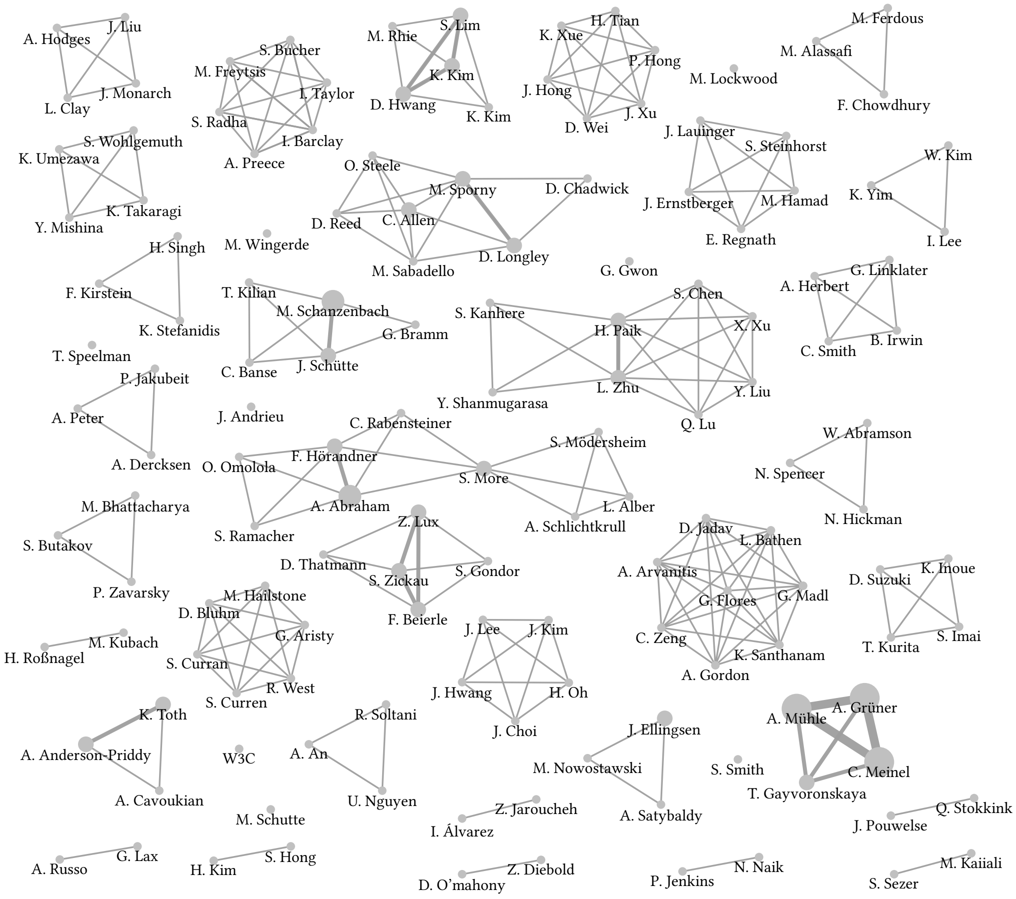


Fig. 6. Co-authorship network graph, where vectors represent authors and edges the co-authorship of one or more works.

The authors with the most publications in this mapping are Andreas Grüner, Alexander Mühle, and Christoph Meinel. They have two articles published together [53, 54] and another two with Tatiana Gayvoronskaya [52, 102]. Therefore, both the vertices and the edges that connect the vertices representing these three authors have the most prominent weight in this graph (*i.e.*, the thickest vertices and edges).

Two researchers have three publications included in this research, Martin Schanzenbach and Andreas Abraham. Schanzenbach's publications include his Ph.D. thesis [115], and two articles with Julian Schütte, one sharing with Georg Bramm [116] and the other with Thomas Kilian and Christian Banse [117]. Abraham's works are a technical report [1], a research paper shared with Felix Hörandner, Olamide Omolola, and Sebastian Ramacher [2], and a third paper also co-authored with Felix Hörandner, which included Christof Rabensteiner and Stefan More [3]. Furthermore, Stefan More published a second paper with other authors [5].

In addition to the aforementioned authors with two mapped publications (Tatiana Gayvoronskaya, Julian Schütte, Felix Hörandner, and Stefan More), there are another seven groups of authors with two publications included in our search: (i) Kalman C. Toth and Alan Anderson-Priddy have published an article together [143] and a second including Ann Cavoukian [144]; (ii) Zoltan A. Lux, Felix Beierle and Sebastian Zickau have worked separately with Sebastian Gondor [92], and with Dirk Thatmann [93]; (iii) Hye Young Paik and Liming Zhu published an article along with Yue Liu, Qinghua Lu, Xiwei Xu, and Shiping Chen [87], and another article with Yashothara Shanmugarasa and Salil S. Kanhere [127]; (iv) Kyung-Hoon Kim, Seungjoo Lim, and Dong-Yeop Hwang published one paper with Ki-Hyung Kim [73] and a second one with Min-Hyung Rhie [82]; (v) Jørgen Ellingsen's master thesis [36] and an article with Abylay Satybaldy, and Mariusz Nowostawski [112]; (vi) Christopher Allen's ten principles [7] and the Decentralized Identifiers (DIDs) standard [109] with Drummond Reed, Ryan Grant, Markus Sabadello, Manu Sporny, and Dave Longley; and (vii) the latter two authors share the verifiable credentials (VC) standard [132] authorship with David Chadwick.

Next, we present the co-reference network of the surveyed literature in Figure 7. In this directed graph, vertices represent publications, and edges represent references between articles, where the destination of an edge means that the source of the edge references this work. The number of received citations defines the vertices' diameter, and the vertices' color is the publishing year.

This chart highlights the importance of Allen's ten principles [7], referenced by thirty-one publications. Likewise, the W3C standards related to SSI, namely DID [109] and VC [132], were the second and third most referenced works with, respectively, twelve and ten references. The first survey of SSI [102], published in 2018, is the fourth most referenced work, with eight mentions.

Regarding cross-references, thirty-seven works are not referenced by any surveyed publication. These thirty-seven unreferenced works are the thirteen publications from 2021, sixteen from eighteen from 2020, seven from thirteen from 2019, and one from five from 2018. Likewise, fifteen publications do not cite any mapped work, in which four are from 2021, two from 2020, four from 2019, one from 2018, three from 2017, and one from 2016. One of the reasons for these fifteen publications without references to other mapped works is the scope of our survey. We excluded SSI platforms such as Sovrin, Uport, and Jolocom, which some of these works reference.

5.3 RQ-2: What conceptual ideas have been introduced or refuted?

Christopher Allen [7] stated that there is no consensus on a definition of SSI, then he proposed ten guiding principles for a starting point, as explained in Section 2. Our second research question aims to map further discussions regarding understanding what SSI is.

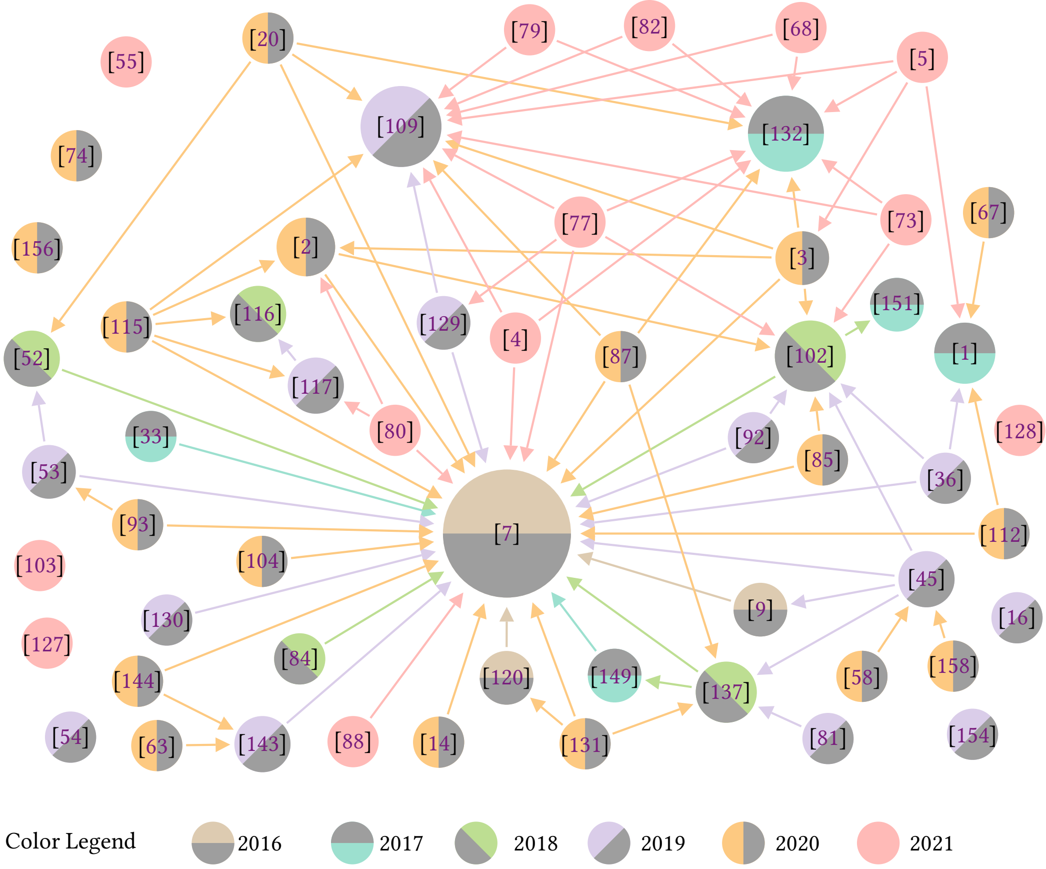


Fig. 7. Co-reference network.

Following our protocol, we mapped fourteen papers that extend Allen's debate about what SSI means. Figure 8 shows this map, which mainly consists of additional specifications/definitions. These are organized into three mutually exclusive subsets of works and five intersecting subsets, one of which groups together papers that refute some of Allen's principles.

The three mutually exclusive subsets are as follows. First, Allen's article makes up a unitary set [7]. Second, three works include the requirement of **verifiable claims** in SSI systems [102, 137, 151]. Third, the subset of articles advocating the need for **GDPR compliance** in SSI systems is also a unitary set [149]. Despite not stating the need for GDPR compliance, the author of [1] debated how SSI systems can use verifiable claims to comply with the following articles of the GDPR: (i) consent; (ii) pseudonymization; (iii) right to erasure; (iv) records of processing activities; (v) data portability; and (vi) data protection by design and by default.

The five subsets of papers with intersecting philosophical matters expose shared ideas of what their authors conceive as mandatory for SSI. The authors of [9, 36, 104] specified that any approach to SSI must be **free** of charge. Furthermore, in addition to the idea of not having to pay to be part of an SSI system, the authors of [9, 104] share with [33] the concept of having the assurance that people's data must be **recoverable** in the event of personal device loss.

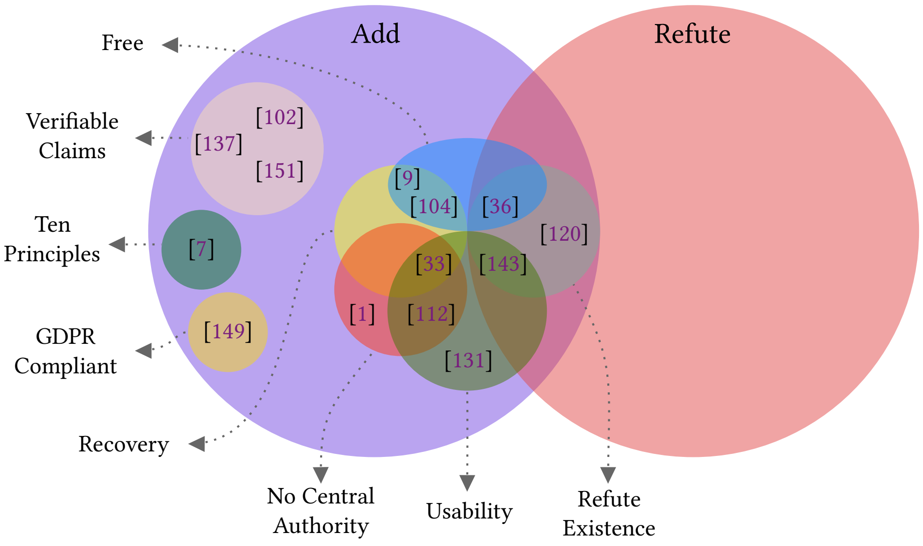


Fig. 8. Map of conceptual works.

Along with [33], both [1] and [112] state that **no central authority** should own or be responsible for an SSI solution. These three articles, along with the papers that advocate that SSI should have no cost [9, 36, 104], have attractive points at first. However, on further inspection, these requirements can make it difficult to develop solutions and build business for SSI. Developers and technicians would have to look for alternative means of revenue and share control over their products. To some extent, this is what Evernym [42], a for-profit organization, did when it spun off Sovrin, a non-profit foundation maintained by various organizations [142]. However, it is not free. Although end users can join the network, receive credentials, and issue verifiable presentations free of charge, companies or other entities that enroll their end users must pay fees to [141]: (i) join the network; (ii) register a credential format, *i.e.*, a credential schema; (iii) begin issuing credentials using a registered schema; (iv) register a revocation registry; and (v) revoke certificates.

Four studies argue that **usability** in SSI is essential [33, 112, 131, 143]. These works declare that users should not need prior knowledge of blockchain technology [33] and other underlying technologies such as cryptographic operations, biometrics, database, and protocols [143]. Moreover, users' needs and expectations must be met and consistent across all platforms and services [112]. One possible approach to achieving these goals could be to mimic physical identities and the interactions we have with them, thus presenting the user to familiar workflows [143]. Ultimately, if the user does not understand what is going on and cannot reason about it, then the individual is not sovereign [131].

Lastly, three works [36, 120, 143] add new ideas and refute concepts introduced by Allen [7]. They all **refute existence**, which states that individuals cannot exist entirely in digital form and (self-sovereign) identities make some aspects of the user public. Toth and Anderson-Priddy [143] have refuted existence, transparency, and protection, arguing that more discussions are needed on these subjects. Similarly, in [36] the authors argued that previous discussions about identity did not address existence [1, 26] and therefore exchanged existence for no cost and unrestricted, stating that an identity system should not impose an economic burden in addition to having a device and connectivity.

Unlike the two studies mentioned above, Schutte [120] discussed Allen's principles through a less technical and more philosophical lens. He argued that an individual, or "self", is not an indivisible entity, but the product of constant interactions by various agents, both internally and externally. He then criticized the principles of existence, control, access, and consent because an individual's identity is not an object but a "heuristic that simplifies information processing and decision making" [120], which is by nature imprecise and therefore it cannot fully anchor identity processes. Ultimately, he argued that claims are of the utmost importance and can be seen as signals published by some actors and perceived by others, who must decide how to prioritize and interpret them.

5.4 RQ-3: What properties or formal definitions have been specified?

The first two years of surveyed publications consist of conceptual works. The works analyzed from 2018 onwards started to present mathematical constructions to describe ideas more accurately. These articles are mapped using our classification mechanism and shown in Figure 9.

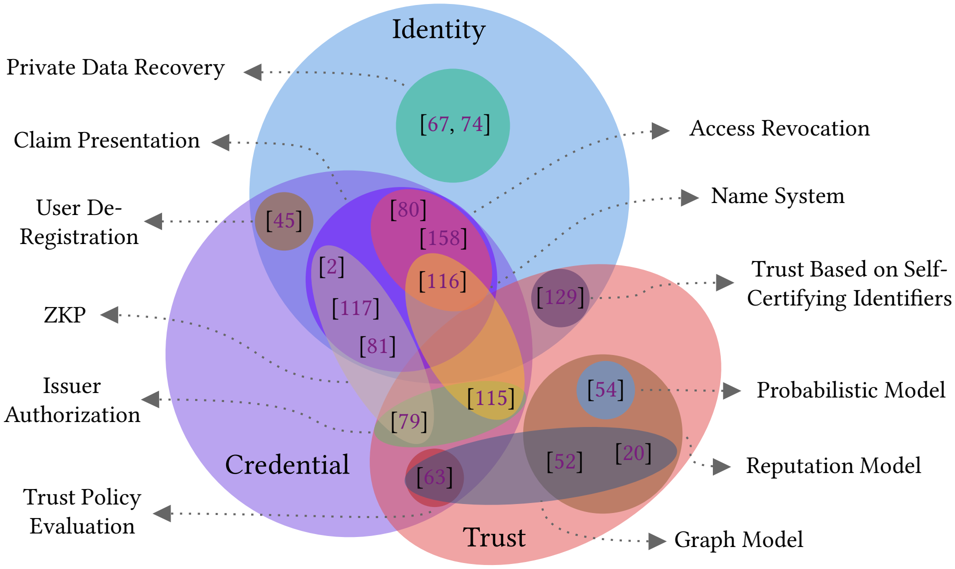


Fig. 9. Map of works that introduce mathematical formalism to SSI.

Following our categorization, two publications are in the identity category [67, 74]. Both papers present mathematical formulations to detail the problem and proposed solutions for **private data recovery**, *e.g.* credentials and claims. In [67], a private data backup system is proposed, in which trusted audit services and trusted individuals are selected. The former receives parts of the keys, while the latter must physically meet to receive encrypted shares of the private data to store on devices with short-range connectivity (such as infrared or near-field communication). Upon losing the private data, trusted peers meet and confirm the newly generated digital identity of the affected user to trusted audit services, which reveal the key needed by the user to decrypt their private data gathered from trusted peers. From a different perspective, [74] uses proxy re-encryption [22]. This technique allows data encrypted with a person's key to be decrypted using someone else's key without revealing anyone's data or key to the proxy. Trusted individuals execute a group key agreement, and then the derived group key is sent to the proxy that contains the encrypted user data. The user's private data can be retrieved from the proxy if the group recreates its key and uses it to authenticate with the proxy, which then uses the proxy re-encryption scheme to have the user's private data accessible to the group.

Three works present **reputation models** [20, 52, 54], **i.e.**, ways to quantitatively assess whether a credential, claim, or identity is trustworthy. Gruner *et al.* [52] built on top of graph theory to create a **graph model** of trust, having a function that searches in the graph for credentials issued to an identity and derives a trust factor. Bhattacharya *et al.* [20] built on top of [52], adding time as a variable in their reputation model. They argued that, in the context of Sovrin, the initial reputation of issuers might come from Sovrin's onboarding process. In contrast, the authors of [54] used probability theory to create a **probabilistic model** of trust. They used probability theory to decide whether claims about the same information, but from different issuers, could be combined to generate trust about it.

Smith [129] focused on **trust based on self-certifying identifiers**. In this work, identifiers are anchored to public-key cryptography and openly reveal the hash of their next public key in their transactions. This preemptive key rotation produces an auditable chain of key transfers of a digital identifier. A distributed ledger is presented as root-of-trust to store the event history of digital identifiers.

With a different perspective, Inoue *et al.* [63] modeled as an Integer Linear Programming (ILP) a **trust policy evaluation** problem specifically for the task of updating a person's information in multiple RPs, each with its trust policy. In this ILP, trust policies are modeled as credibility requirements about incoming claims, and successfully updating a person's information in an RP increases credibility. The ILP is then transformed into a graph problem, and a heuristic based on Dijkstra's algorithm finds approximate solutions. This article is the only surveyed material with a formal description of the problem attacked.

Two works present ideas and mathematical models to implement **issuer authorization** [79, 115], which allows issuers to establish hierarchies similar to traditional certificate authorities in PKI. Schanzenbach's Ph.D. thesis [115] presents a construction using name systems (*e.g.* Domain Name System (DNS) [99], and GNU Name System (GNS) [152]) that allows an issuer to delegate authorization to issue credentials with specific attributes to other issuers, which can also delegate to others. With the same goal, but using a different construction, the authors of [79] formalized a model that uses the RSA cryptographic accumulator [25] to allow authorized issuers to issue credentials without revealing their identity. The authors argued that this fills a gap in the Hyperledger Indy framework [60], where an issuer A cannot stop B from issuing credentials using A's credential format.

In addition to being part of the group of works that model issuer authorization, [79] is also present in the group of works that formalize **Zero-Knowledge Proof (ZKP)** techniques. In this work, ZKP allows an issuer to mathematically prove that it is part of a group of authorized issuers without revealing a unique identifier. The other three works in this group use ZKP to do **claim presentation** (*i.e.* verifiable presentation). Both [81] and [117] use a general-purpose ZKP named zero-knowledge Succinct Non-interactive ARguments of Knowledge (zk-SNARK) [21] to allow the subject of the credentials to produce proofs about the data in their credentials. Through zk-SNARK, proofs containing predicates over attributes in credentials are possible, *e.g.* the credential subject is over 18 years old. In contrast, Abraham *et al.* [2] built a ZKP proof system using Water's signature [153] and BLS signature [23] that does not allow the execution of predicates over certificate data. Instead, their scheme allows users to reveal certificate attributes without showing the entire certificate.

Three other works also model claim presentation, and additionally, introduce **access revocation** of presented claims [80, 116, 158]. In [80] a model is detailed where social media platforms such as Facebook and LinkedIn are used as a means of requesting, generating, and revoking credentials, along with claim presentation and revocation of presented claims. However, predicates over credential attributes are not supported, only attribute disclosure. Differently, the authors of [158] implemented claims revocation through chameleon hashing [76]. This family of one-way functions uses a trapdoor so that, without access to it, they function like traditional one-way functions. However, having access to the trapdoor, *e.g.*, through a key, one can easily find collisions for a given input. [12, 31] used this special feature of the chameleon hash to implement a rewriteable blockchain, that is, a blockchain whose history can be manipulated through the chameleon hash trapdoor. Based on these efforts, the authors of [158] implemented their blockchain in which users can revoke access to claims published in the ledger through a trapdoor. Lastly, in [116], it is argued that static claims that a specific party has been granted access cannot be revoked because it is likely that they have been persisted locally. Their approach to this challenge is to grant and revoke access to up-to-date information through version control and encryption. Keys are provided to RPs as new versions of claims are created, thereby revoking access as needed.

Finally, a unit set is composed of [45], where Ferdous *et al.* provided a comprehensive mathematical model of SSI. This formalization includes user registration and **user de-registration**.

5.5 RQ-4: What practical problems have been introduced and solved?

Our keywording process has enabled us to classify surveyed works and generate visualizations to answer our research questions. The data items on our data extraction form relating to our last research question, namely *What practical problems have been introduced and solved?*, have been filled in by most of the mapped works. Consequently, the map for our fourth research question is denser, having a larger number of publications when compared to the previous ones. Figure 10 shows this map.

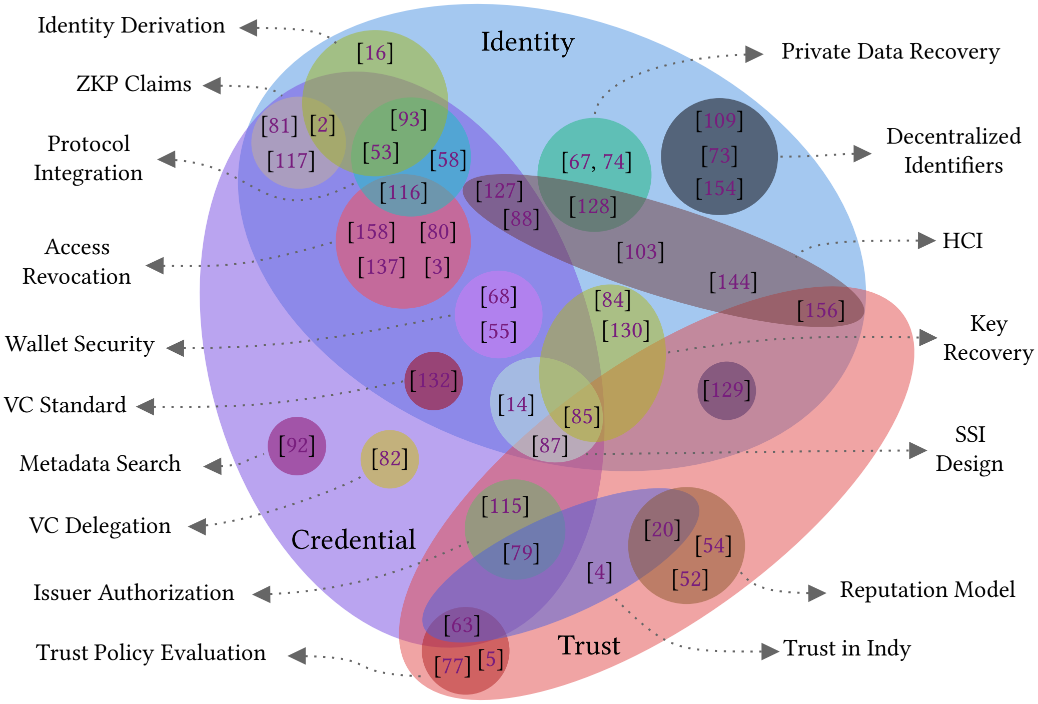


Fig. 10. Map of publications that introduced and solved novel problems in the SSI ecosystem.

In the identity set there are two subsets disjoint to the credential and trust sets, namely **decentralized identifiers** and **private data recovery**. The former holds two standards [109, 154] and one research article [73].

In [109] the Decentralized Identifiers (DID) standard is proposed. It introduces a metamodel for creating identifiers without centralized hierarchy and controlled by their owners. An instance of this metamodel is a DID method and defines specific details such as underlying encryption algorithms and how the identifiers in this method are unique. All DIDs are Uniform Resource Identifiers (URIs) [18] with three parts separated by a colon: (i) the `did` scheme identifier; (ii) the DID method identifier; and (iii) the DID method-specific identifier. For instance, `did:key:z6MkpTHR8VNsBxYAAWHut2Geadd9jSwuBV8xRoAnwWsdvktH` is a valid DID identifier that uses the DID method named `key` [134]. In this method, the first character of the method-specific identifier is always `z`, and the following three characters represent the public-key algorithm used. In this case, the characters `6Mk` indicate that `Ed25519` [70] was used, and subsequent characters are the multibase [17] encoded public-key. Other DID methods use blockchain and other technologies to persist the user-generated DID and its respective DID document, a JSON-based document with communication endpoints and cryptographic keys to ensure that the holder of a DID is its owner. As pointed out in Section 5.2.3, this standard has significant importance in the SSI literature, being the second most referenced work in this survey.

Although [109] establishes a starting point for self-sovereign identifiers and the authentication of their owners, this standard does not specify a way for two (or more) DIDs to interact. [154] attacks this problem proposing DIDComm, a two-party protocol to create a communication channel between the holders of two DIDs with pre-defined messages, states, and transitions. It supports the transportation of messages through out-of-band channels such as QRcode and e-mail [155].

According to Kim *et al.* [73], the endpoint URLs in DID documents have an anonymity problem. They argued that URLs could leak information, such as the country of origin and other affiliations. They proposed two countermeasures: (i) to suppress URLs altogether and use other means of communication; and (ii) to use gateway URLs that only redirect authorized entities to the correct address.

With regards to **private data recovery**, there are three works that present this problem and propose solutions [67, 74, 128]. Two of them [67, 74] introduce mathematical formalism and therefore were detailed above. The remaining article [128] in this category also concerns **human-computer interaction (HCI)**. The authors argued that their scheme presents a trade-off between security, *i.e.*, storing an encrypted form of the private key in lower security environments, and usability, *i.e.*, recovering the original private key without the need for long passphrases or hardware security modules (HSMs). To achieve this trade-off, the private key is divided using Shamir's secret sharing technique [126], a secret sharing protocol where the secret is recovered if a minimum number of parts can be retrieved. In this case, the user has to correctly answer a minimum amount of previously registered questions, where each answer is a part of Shamir's technique. To increase security, the minimum amount of correct answers is increased.

Five other articles also present concerns about HCI in SSI. Two are in the identity category [103, 144], two are in the intersection between identity and credential [88, 127], and one is in the intersection between identity and trust [156].

Toth *et al.* [144] argued that biometrics and other forms of two-factor authentication slightly improved the security of identities. Then, they introduced a software agent to manage user data. It helps users decide which identities to create and use and which private information to disclose, thereby improving security through better human-computer interactions. With a different focus, the authors of [103] put forward a patent for an authentication method based on a person's interactions with its personal device. The device monitors application usage patterns, browser history, location history, and other measurements to determine if the person holding the device is the owner.

Regards HCI and trust, [156] suggests that the problem of deciding whether or not to trust an identity and its claims is a substantial risk for an algorithm to decide on its own. They put forward a proposal where the user has to decide if digital identities can be trusted or not actively. The user is empowered to make that decision through a graph of previous interactions of the proponent with other digital identities, which is created from the history stored in a distributed ledger.

The last two works in the HCI subset are at the intersection of identity and credentials [88, 127]. The authors of [88] presented an extensive study of SSI usability and found out that the current interactions of SSI systems require extensive prior knowledge and participant responsibility. The authors used the human data interaction theory [101], which argues that humans interact with data rather than computers, to investigate the interface layer of SSI. The conclusion points to the need for standardization and design thinking of interfaces and interactions to increase the likelihood of adoption. On the subject of participant responsibility, Shanmugarasa *et al.* [127] tackled the problem of users managing verifiable presentations. For example, non-technically savvy users may accept to provide more information than requesting services need. The proposed solution for this problem is a privacy preference recommendation system that uses machine learning algorithms with pre-trained models built from surveyed data on privacy preferences. This system helps the user by advising which attributes are acceptable to be shared.

Five works present the problem of **access revocation** of credentials [3] or presented claims [80, 116, 137, 158]. Three of them [80, 116, 158] use mathematical formalism, and therefore have been described above. The remaining two articles [3, 137] are detailed next.

Concerned about the portability and interoperability of claims, the authors of [137] introduced a metamodel for specifying claims in blockchains. First, they described the claim metadata composed of name, timestamp, expiration time, proof format, and proof link. Then the claims lifecycle is organized in a blockchain format, where there are specific types of blocks for making new claims, adding more attestation to a claim (*i.e.*, having other entities sign a claim), and revoking an attestation or a claim. On the other hand, Abraham *et al.* [3] attacked the problem of offline verification of credential status. Their approach is to have the blockchain generate an attestation of the validity of requested certificates with the timestamp. This attestation is then presented when there is no connectivity to the revocation registry, and the relying party decides if it is recent enough to be accepted.

Another trend of research in SSI is to provide **protocol integration** with production-level standards. This was presented as the driving problem of [53, 58, 93], but it was also developed in the aforementioned work of Schanzenbach *et al.* [116]. Both [53] and [93] aim to integrate SSI with the OpenID Connect protocol [111], an identity layer that provides authentication on top of OAuth 2.0, the *de facto* industry-standard protocol for authorization [57].

In [53], a gateway is built between two SSI solutions [43, 146] and web applications using the OpenID Connect protocol. Users can select claims to compose their identity, which are verified by the gateway and then transferred to the destination application for authentication using the OpenID Connect protocol. Similarly, [93] implements a gateway between Hyperledger Indy [60] and other applications through OpenID Connect, where users of any instance of Hyperledger Indy (such as Sovrin [140]) can benefit. Unlike [53], a wallet application is developed to store credentials on the user's smartphone. Application-level authorization is implemented in [93] with claims, which the user must present. Differently, the authors of [58] implemented authorization through OAuth 2.0, which facilitates integration with existing web services. However, unlike [53, 93], authentication in [58] uses a custom mechanism instead of OpenID Connect.

By allowing users of SSI solutions to access web applications through the OpenID Connect protocol, [53, 93] ended up implementing **identity derivation** mechanisms, that is, a way to derive an SSI identity from non-SSI systems. This is the main objective of [2] but also achieved by [16]. In [2], a digital identity derivation protocol is proposed, where user data scattered across various IdPs are gathered and transformed into credentials to be imported into an SSI system. It starts by connecting the individual to the IdPs that hold the data, and then the attributes are signed. The attributes in these credentials can be presented as verifiable claims through ZKP. Focusing on digital identity authentication, Bathen *et al.* [16] discussed the possibility of replay attacks using biometric data once an attacker gains access to biometric templates. They argued that the solution to this problem is user-managed cancelable biometrics. In this system, a person's self-image, *i.e.* a selfie, is passed through one-way functions to mask the original data, then the result is stored on a blockchain and managed as a credential. This second authentication factor can be used for a higher level of assurance authentication.

Along with [2], two other surveyed articles propose zero-knowledge proof schemes to create provable claims [81, 117], labeled **ZKP claims** in Figure 10. Those three papers introduced mathematical formalism and were described above.

The subject of **wallet security** drives one patent [55] and one research article [68]. The authors of [55] proposed a hardware-based wallet containing cryptographic keys and credentials. It can be connected to mobile devices when needed and disconnected, and safely stored after usage. On the other hand, in [68] holders do not store their credentials. Instead, credentials are hosted on a storage service and secured by a two-party protocol. Furthermore, holders do not access their data directly. Instead, RP agents request the holder's agent running on the storage service for the needed information, which then asks the user for authentication and authorization. With this approach, users never receive their credentials, and therefore do not have to worry about storing them securely. As a two-party encryption protocol encodes the credentials, the storage service cannot misuse the credentials.

The most influential work in the credential set, according to our findings above, is the W3C's **Verifiable Credentials standard** [132]. Like the DID standard [109], this standard is also a metamodel. It defines the meta-structure and lifecycle of VCs and VPs. Both VC and VP must have: (i) metadata that describes the data; (ii) the data; and (iii) cryptographic proof of integrity and authenticity. When some model instantiates this metamodel, it needs to specify the adopted syntax, cryptographic algorithms, and proof format to construct VCs and VPs. For instance, in Hyperledger Indy [60], a VC's metadata is stored in a distributed ledger, while the data and proof are stored in a JSON file.

There are two research materials [82, 92] in the credential set that are neither intersecting with identity nor trust. The authors of [92] introduced the problem of **metadata search** in Hyperledger Indy [60]. The authors argued that adding new types of credentials comes at a monetary cost, and therefore it is worthwhile to reuse existing formats. They used Apache Solr [10] to build a search application that allows users to find existing credential metadata stored on the ledger. Attacking a different challenge, Lim *et al.* [82] proposed a **VC delegation** mechanism that requires the VC subject to confirm or deny its usage by the delegatee. They argued that a VP constructed by delegatees is limited because they only have the VC in an encrypted format. Hence, any VP presented by a delegatee incurs in communication with the VC subject to request authorization and complement the VP with the required data.

Three articles discuss issues related to **SSI Design** [14, 85, 87], where [85, 87] are the only works at the intersection of the three major categories of our map. In [87] design patterns are presented to facilitate the development of new blockchain-based SSI applications. The lifecycle of key management, identity management, and credential management are presented. Then twelve patterns are proposed within these three groups, following the format of [94], which consists of a pattern name, summary, the context of use, a problem statement, a discussion, the solution, and its consequences. Differently, the authors of [85] argued that IDMs could be reduced to two mappings: (i) digital identifier and its owner; and (ii) digital identifier and its credentials. In addition, the following operations are required for both mappings: create, read, update, delete and verify. The way they are built depends on the desired trust model for the system. If SSI is the goal, all of them should be done without relying on any authority. Finally, Barclay *et al.* [14] proposed a modeling strategy in which SSI entities and their credentials can be specified and understood by non-technical stakeholders. They used iStar 2.0 [30], an actor-based modeling language that allows for the representation of actors and the interdependence of their goals, to represent users in the issue of credentials and presentation of claims in an SSI system.

One of the design patterns specified in [87] concerns **key recovery**. Two articles focus mainly on this challenge [84, 130]. In [84], a self-signed root certificate acts as a certificate authority (CA) that generates short-lived intermediate certificates for the users. The authors argued that since certificates are rotated at a predetermined frequency, the key recovery problem ceases to exist as long as the root certificate and its private key are not lost or compromised. In contrast, Soltani *et al.* [130] tackled the key recovery problem through a decentralized protocol. They developed a wallet application where the users pre-picks their trusted peers and which keys will be recoverable. Key pieces are distributed to pre-selected users in a secret sharing protocol based on [126] and can be recovered by its owner if a minimum number of parts can be retrieved from peers.

Three research papers attack the problem of **trust policy evaluation** [5, 63, 77], that is, how to define trust rules and relationships so that a machine can automatically evaluate them. This challenge is not the main focus of [63] as explained above. Nonetheless, an ILP model and a graph-based heuristic are introduced. The authors of [77] proposed that entities define trust policies through lists of authorities they trust. These trusted entities, in turn, also publish which entities they recognize as trustworthy. For instance, one could trust a bank federation that periodically reports which banks it recognizes as credible. Thus, when receiving the VP of a person stating that she has an account on an unrecognized bank, a query to the bank federation's list of trusted banks is enough to decide if the VP can be trusted or not. In contrast, in [5] the trust policy language (TPL) [100], a declarative language for specifying trust rules without worrying about low-level aspects, was adapted to work in SSI. The TPL has been extended with SSI-related concepts such as DID and VC, allowing for the specification of rules to check VPs.

The two works [79, 115] mapped relating to the **issuer authorization** introduced mathematical formalism and were detailed above. Nevertheless, [79] shares with two aforementioned papers [20, 63] and [4] the characteristic of basing their approach to trust in Hyperledger Indy. More specifically, [4] presented the different user roles and transaction types stored in the Indy blockchain and then discussed the steps a verifier can take to gain confidence when receiving a presentation. For instance, they argued that if different entities issue credentials of a given format (credential schema), this provides more assurance than a schema that a single issuer only endorses.

Lastly, [20, 52, 54] form the set of works that deal with **reputation model** in SSI. This set of articles, along with the unit set of **trust based on self-certifying identifiers** [129], whose label is omitted in Figure 10 for brevity, were discussed earlier. Thus, we categorized and detailed all surveyed works that advance the state of the art of SSI. In the next section, we proceed to discuss open challenges.

6 OPEN CHALLENGES

The surveyed materials present initial developments on SSI. New publications will continue to advance the conceptual debate about what it means for an identity to be self-sovereign and introduce new and unforeseen challenges in the SSI ecosystem. Based on the evidence gathered to answer our research questions, we identify challenges that should be attacked by future work. They are detailed below with recommendations.

A definition of SSI that researchers and practitioners accept. In this review, we have gathered evidence (see Section 5.3) that most of the articles reviewed on SSI fundamentals agree with the principles established by Allen [7], also adding new ones. Promoting a comprehensive review and discussion is important to produce a new set of rules to define SSI. In addition, mathematical formalism can be used to establish precise boundaries. Having an accurate definition of SSI will benefit future efforts and, ultimately, users who will transition between SSI systems knowing they share the same fundamentals.

Basic research. Most of the materials surveyed provide a mathematical model tailored to the proposed algorithm. Only one of the articles surveyed presents a comprehensive mathematical formulation of SSI [45], but it does not cover the inherently decentralized trust aspects of SSI. In addition, [87] makes realistic considerations and provides design patterns for many facets of SSI, including trust. These articles [45, 87] are good starting points. However, more basic research is needed to promote a debate about ideas on how to jointly represent identities, credentials, claims, and trust, which is important for future pragmatic research. By answering RQ-2 and RQ-3 (see Sections 5.3 and 5.4), we produce a starting point for future basic research.

Complex attribute sharing. Revised publications: (i) use ZKP to create Boolean predicates about attributes [2, 81, 117]; (ii) create versions of the data, encrypting and distributing keys to the most recent revision [116]; or (iii) use a chameleon hash to implement redactable blockchains [158]. However, these methods are unsuitable when more complex attributes need to be shared and may not change for several years. For instance, the shipping address for online purchase or a phone number. Therefore, more research on VP is required to ensure a wide variety of use cases coverage.

Sound trust models. Trust plays an essential role in SSI and will be of paramount importance for the adoption of SSI solutions. If trust models are not tested comprehensively, they will become attractive targets for hackers. This challenge is exacerbated by the current standardization effort [133] that specifies a Boolean trust model in which a verifier trusts the issuer or not, which does not cover fuzzy scenarios of the real world. For example, an entity may present multiple claims about the same attribute where some issuers are trusted, and others are not. Can this claim be trusted? Quantifiable trust/reputation models are needed, and only three articles surveyed attack this problem [20, 52, 54]. Furthermore, trust models need solid security, so formal verification techniques [95] must be employed to provide an adequate level of assurance.

Blockchainless SSI. On blockchain-based SSI systems, trust in an IdP has not been completely removed but instead replaced by a decentralized entity, which the user needs to trust to embrace SSI. The user should not have to trust and depend on a blockchain consortium to participate in an SSI ecosystem. However, most publications work with the flawed assumption that blockchain is a fundamental part of SSI. The user should not need to trust anyone to be self-sovereign.

To facilitate the migration from other paradigms. In federated and user-centric models, the administration's burden falls almost entirely in the IdP. Users usually need to just worry about password management. With SSI, users are also overloaded with management tasks such as backing up their keys, identities, and credentials and creating and presenting claims. We have mapped publications that propose techniques to derive (self-sovereign) identities from federated and user-centric identities [2, 16], and others that worry with the backup and recovery of keys [84, 130] and private data [67, 74]. Therefore, academia is picking up momentum on this migration challenge.

Usability. It is important to research applications' interfaces and the interactions between people through them. Ultimately, individuals will use SSI systems. Meaningful interaction must occur between people and applications and, more importantly, between individuals in an SSI ecosystem. Otherwise, users are unlikely to leave the comfort of their current federated/user-centric identities. A common trend in usability research is to imitate physical wallets [131, 144], thus presenting the user with everyday interactions. Innovative solutions are necessary and can be decisive for the widespread adoption and success of SSI.

7 FINAL REMARKS

SSI is a new identity management paradigm that increases people's agency in the digital world. Boosted by the popularity of blockchain, SSI systems attracted attention from academia and industry. Existing surveys have presented biased results towards blockchain and missed the bigger picture.

In this article, we systematically surveyed both peer-reviewed literature and non-peer-reviewed literature that: (i) expanded the conceptual discussion on what SSI is; (ii) introduced mathematical formulation to precisely define one or more SSI-related problems; or (iii) introduced a novel pragmatical problem related to the SSI ecosystem and presented a solution to it. After keywording the selected works, three SSI-specific maps were created and used to classify the surveyed work. These maps allow the reader to understand the main contributions of the reviewed works and give a broad understanding of the current position of research in SSI. In addition, our maps serve as a basis for researchers and entrepreneurs who intend to expand SSI conceptually or develop new SSI systems. Finally, we presented open challenges, thus leading to future research and opportunities directions in the area.

ACKNOWLEDGMENTS

This study was supported by the Federal Institute of Education, Science and Technology of Rio Grande do Sul (IFRS).

REFERENCES

[1] Andreas Abraham. 2017. *Self-sovereign identity: Whitepaper about the Concept of Self-Sovereign Identity including its Potential.* Technical Report. A-SIT. Retrieved 2021-08-03 from https://technology.a-sit.at/en/whitepaper-self-sovereign-identity/

[2] Andreas Abraham, Felix Hörandner, Olamide Omolola, and Sebastian Ramacher. 2020. Privacy-Preserving eID Derivation for Self-Sovereign Identity Systems. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 11999 LNCS. Springer International Publishing, Copenhagen, Denmark, 307–323. https://doi.org/10.1007/978-3-030-41579-2_18

[3] Andreas Abraham, Stefan More, Christof Rabensteiner, and Felix Hörandner. 2020. Revocable and offline-verifiable self-sovereign identities. In *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. IEEE, IEEE Computer Society, Guangzhou, China, 1020–1027. https://doi.org/10.1109/TrustCom50675.2020.00136

[4] Will Abramson, Nicky Hickman, and Nick Spencer. 2021. Evaluating Trust Assurance in Indy-Based Identity Networks Using Public Ledger Data. *Frontiers in Blockchain* 4 (2021), 18. https://doi.org/10.3389/fbloc.2021.622090

[5] Lukas Alber, Stefan More, Sebastian Mödersheim, and Anders Schlichtkrull. 2021. Adapting the TPL Trust Policy Language for a Self-Sovereign Identity World. In *Open Identity Summit 2021*, Heiko Roßnagel, Christian H. Schunck, and Sebastian Mödersheim (Eds.). Gesellschaft für Informatik e.V., Lyngby, Denmark, 107–118. Retrieved 2021-08-03 from https://dl.gi.de/handle/20.500.12116/36506

[6] Sinică Alboaie and Doina Cosovan. 2017. Private Data System Enabling Self-Sovereign Storage Managed by Executable Choreographies. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 10320 LNCS. Springer Verlag, Neuchâtel, Switzerland, 83–98. https://doi.org/10.1007/978-3-319-59665-5_6

[7] Christopher Allen. 2016. *The path to self-sovereign identity.* Life with Alacrity. Retrieved 2021-01-14 from http://www.lifewithalacrity.com/2016/04/the-path-to-self-soverereign-identity.html

[8] Gergely Alpár, Fabian van den Broek, Brinda Hampiholi, Bart Jacobs, Wouter Lueks, and Sietse Ringers. 2017. IRMA: practical, decentralized and privacy-friendly identity management using smartphones. In *10th Workshop on Hot Topics in Privacy Enhancing Technologies (HotPETs 2017)*. HotPETs, Minneapolis, USA, 1–2. Retrieved 2021-01-14 from https://www.petsymposium.org/2017/papers/hotpets/irma-hotpets.pdf

[9] Joe Andrieu. 2016. A Technology-Free Definition of Self-Sovereign Identity. In *Rebooting the Web of Trust III*. Web of Trust, San Francisco, USA, 2–5. Retrieved 2021-01-14 from https://github.com/WebOfTrustInfo/rwot3-sf/blob/master/topics-and-advance-readings/a-technology-free-definition-of-self-sovereign-identity.pdf

[10] Apache Software Foundation. 2010. Apache Solr. Retrieved 2021-01-14 from https://lucene.apache.org/solr/

[11] Association for Computing Machinery. 2010. ACM Digital Library. Retrieved 2021-01-14 from https://dl.acm.org

[12] Giuseppe Ateniese, Bernardo Magri, Daniele Venturi, and Ewerton Andrade. 2017. Redactable Blockchain - or - Rewriting History in Bitcoin and Friends. In *2017 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, IEEE Computer Society, Paris, France, 111–126.

https://doi.org/10.1109/EuroSP.2017.37

[13] World Bank. 2019. ID4D Practitioner's Guide. Retrieved 2021-08-16 from http://documents.worldbank.org/curated/en/248371559325561562/ID4D-Practitioner-s-Guide

[14] Iain Barclay, Maria Freytsis, Sherri Bucher, Swapna Radha, Alun Preece, and Ian Taylor. 2020. Towards a Modelling Framework for Self-Sovereign Identity Systems. *arXiv e-prints* abs/2009.04327 (9 2020), 1–5. arXiv:2009.04327 Retrieved 2021-01-14 from https://arxiv.org/abs/2009.04327

[15] Mathieu Bastian, Sebastien Heymann, Mathieu Jacomy, et al. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. *The International AAAI Conference on Web and Social Media (ICWSM)* 8, 2009 (2009), 361–362. Retrieved 2021-01-14 from https://gephi.org/publications/gephi-bastian-feb09.pdf

[16] Luis Bathen, German H. Flores, Gabor Madl, DIvyesh Jadav, Andreas Arvanitis, Krishna Santhanam, Connie Zeng, and Alan Gordon. 2019. SelfIs: Self-sovereign biometric IDs. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE Computer Society, Long Beach, USA, 2847–2856. https://doi.org/10.1109/CVPRW.2019.00344

[17] Juan Benet and Manu Sporny. 2021. *The Multibase Data Format*. Internet-Draft draft-multiformats-multibase-03. Internet Engineering Task Force. https://datatracker.ietf.org/doc/html/draft-multiformats-multibase-03 Work in Progress.

[18] Tim Berners-Lee, Roy T. Fielding, and Larry M Masinter. 2005. Uniform Resource Identifier (URI): Generic Syntax. RFC 3986. https://doi.org/10.17487/RFC3986

[19] Elisa Bertino and Kenji Takahashi. 2010. *Identity management: Concepts, technologies, and systems*. Artech House, London, United Kingdom.

[20] Manas Pratim Bhattacharya, Pavol Zavarsky, and Sergey Butakov. 2020. Enhancing the Security and Privacy of Self-Sovereign Identities on Hyperledger Indy Blockchain. In *2020 International Symposium on Networks, Computers and Communications (ISNCC)*. IEEE, IEEE Computer Society, Montreal, Canada, 1–7. https://doi.org/10.1109/ISNCC49221.2020.9297357

[21] Nir Bitansky, Ran Canetti, Alessandro Chiesa, and Eran Tromer. 2012. From extractable collision resistance to succinct non-interactive arguments of knowledge, and back again. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*. Association for Computing Machinery, New York, USA, 326–349. https://doi.org/10.1145/2090236.2090263

[22] Matt Blaze, Gerrit Bleumer, and Martin Strauss. 1998. Divertible protocols and atomic proxy cryptography. In *International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, Springer, Espoo, Finland, 127–144. https://doi.org/10.1007/BFb0054122

[23] Dan Boneh, Ben Lynn, and Hovav Shacham. 2001. Short signatures from the Weil pairing. In *International conference on the theory and application of cryptology and information security*. Springer, Springer, Gold Coast, Australia, 514–532. https://doi.org/10.1007/3-540-45682-1_30

[24] Jan Camenisch and Anna Lysyanskaya. 2001. An efficient system for non-transferable anonymous credentials with optional anonymity revocation. In *Advances in Cryptology — EUROCRYPT 2001*, Birgit Pfitzmann (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 93–118. https://doi.org/10.1007/3-540-44987-6_7

[25] Jan Camenisch and Anna Lysyanskaya. 2002. Dynamic Accumulators and Application to Efficient Revocation of Anonymous Credentials. In *Advances in Cryptology — CRYPTO 2002*, Moti Yung (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 61–76. https://doi.org/10.1007/3-540-45708-9_5

[26] Kim Cameron. 2005. *The Laws of Identity*. Technical Report. Microsoft Corporation. Retrieved 2021-01-14 from https://www.identityblog.com/stories/2005/05/13/TheLawsOfIdentity.pdf

[27] Edwin Pun Chau and Robert Hertzberg. 2018. *California Consumer Privacy Act of 2018*. California State Legislature. Retrieved 2021-01-10 from https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375

[28] David Chaum. 1985. Security without Identification: Transaction Systems to Make Big Brother Obsolete. *Commun. ACM* 28, 10 (Oct. 1985), 1030–1044. https://doi.org/10.1145/4372.4373

[29] Clarivate Analytics. 2006. Web of Science. Retrieved 2021-01-14 from https://www.webofknowledge.com/

[30] Fabiano Dalpiaz, Xavier Franch, and Jennifer Horkoff. 2016. iStar 2.0 Language Guide. *arXiv* abs/1605.07767 (2016), 1–15. Retrieved 2021-01-14 from https://arxiv.org/abs/1605.07767

[31] David Derler, Kai Samelin, Daniel Slamanig, and Christoph Striecks. 2019. Fine-Grained and Controlled Rewriting in Blockchains: Chameleon-Hashing Gone Attribute-Based. Cryptology ePrint Archive, Report 2019/406. Retrieved 2021-01-14 from https://eprint.iacr.org/2019/406

[32] Omar Dib and Khalifa Toumi. 2020. Decentralized identity systems: Architecture, challenges, solutions and future directions. *Annals of Emerging Technologies in Computing (AETiC), Print ISSN* 4, 5 (2020), 2516–0281. https://doi.org/10.33166/AETiC.2020.05.002

[33] Zachary Diebold and Donal O'Mahony. 2017. *Self-Sovereign Identity using Smart Contracts on the Ethereum Blockchain*. Master's thesis. University of Dublin. Retrieved 2021-01-14 from https://www.scss.tcd.ie/publications/theses/diss/2017/TCD-SCSS-DISSERTATION-2017-016.pdf

[34] eBay Inc. 1995. eBay. Retrieved 2021-08-17 from https://www.ebay.com/

[35] Tewfiq El Maliki and Jean-Marc Seigneur. 2007. A Survey of User-centric Identity Management Technologies. In *The International Conference on Emerging Security Information, Systems, and Technologies (SECUREWARE 2007)*. IEEE, IEEE Computer Society, Valencia, Spain, 12–17. https://doi.org/10.1109/SECUREWARE.2007.4385303

[36] Jørgen Ellingsen. 2019. *Self-Sovereign Identity Systems Opportunities and challenges*. Master's thesis. Norwegian University of Science and Technology. Retrieved 2021-01-14 from https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2617756

[37] Elsevier. 1997. ScienceDirect. Retrieved 2021-01-14 from https://www.sciencedirect.com/

[38] Elsevier. 2008. Mendeley. Retrieved 2021-01-14 from https://www.mendeley.com/

[39] Elsevier. 2010. Scopus Preview. Retrieved 2021-01-14 from https://www.scopus.com/

[40] Emir Erdem and Mehmet Tahir Sandıkkaya. 2018. OTPaaS—One time password as a service. *IEEE Transactions on Information Forensics and Security* 14, 3 (2018), 743–756. https://doi.org/10.1109/TIFS.2018.2866025

[41] European Parliament, Council of the European Union. 2016. *Regulation (EU) 2016/679*. European Parliament. Retrieved 2021-01-10 from http://data.europa.eu/eli/reg/2016/679/oj

[42] Evernym. 2013. Evernym. Retrieved 2021-01-14 from https://www.evernym.com/

[43] Ch Fei, J Lohkamp, E Rusu, K Szawan, K Wagner, and N Wittenberg. 2018. *Self-Sovereign and Decentralised Identity By Design*. Technical Report. Jolocom. Retrieved 2021-01-14 from https://github.com/jolocom/jolocom-lib/wiki/Jolocom-Whitepaper

[44] Md Ferdous et al. 2015. *User-controlled Identity Management Systems using mobile devices*. Ph.D. Dissertation. University of Glasgow. Retrieved 2021-01-14 from http://theses.gla.ac.uk/6621/

[45] Md Sadek Ferdous, Farida Chowdhury, and Madini O. Alassafi. 2019. In Search of Self-Sovereign Identity Leveraging Blockchain Technology. *IEEE Access* 7 (2019), 103059–103079. https://doi.org/10.1109/ACCESS.2019.2931173

[46] Md. Sadek Ferdous, Gethin Norman, and Ron Poet. 2014. Mathematical Modelling of Identity, Identity Management and Other Related Topics. In *Proceedings of the 7th International Conference on Security of Information and Networks* (Glasgow, Scotland, UK) *(SIN '14)*. Association for Computing Machinery, New York, NY, USA, 9–16. https://doi.org/10.1145/2659651.2659729

[47] Eva Galperin and Wafa Ben Hassine. 2015. *Changes to Facebook's "Real Names" Policy Still Don't Fix the Problem.* Electronic Frontier Foundation. Retrieved 2021-01-14 from https://www.eff.org/deeplinks/2015/12/changes-facebooks-real-names-policy-still-dont-fix-problem

[48] Komal Gilani, Emmanuel Bertin, Julien Hatin, and Noel Crespi. 2020. A survey on blockchain-based identity management and decentralized privacy for personal data. In *2020 2nd Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS)*. IEEE, IEEE, Paris, France, 97–101. https://doi.org/10.1109/BRAINS49436.2020.9223312

[49] Google. 2004. Google Scholar. Retrieved 2021-01-14 from https://scholar.google.com/

[50] Google. 2006. Google Patents. Retrieved 2021-07-22 from https://patents.google.com/

[51] UK government digital service. 2014. Introducing GOV.UK Verify. Retrieved 2021-08-16 from https://www.gov.uk/government/publications/introducing-govuk-verify/introducing-govuk-verify

[52] Andreas Grüner, Alexander Mühle, Tatiana Gayvoronskaya, and Christoph Meinel. 2018. A Quantifiable Trust Model for Blockchain-Based Identity Management. In *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. IEEE, Halifax, NS, Canada, 1475–1482. https://doi.org/10.1109/Cybermatics_2018.2018.00250

[53] Andreas Grüner, Alexander Mühle, and Christoph Meinel. 2019. An Integration Architecture to Enable Service Providers for Self-sovereign Identity. In *2019 IEEE 18th International Symposium on Network Computing and Applications, NCA 2019*. IEEE, Cambridge, USA, 261–265. https://doi.org/10.1109/NCA.2019.8935015

[54] Andreas Gruner, Alexander Muhle, and Christoph Meinel. 2019. Using Probabilistic Attribute Aggregation for Increasing Trust in Attribute Assurance. In *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, Xiamen, China, 633–640. https://doi.org/10.1109/SSCI44817.2019.9003094

[55] Oh Gyoung GWON. 2021. Content wallet device and self-sovereign identity and copyright authentication system using same. Retrieved 2021-07-16 from https://patents.google.com/patent/WO2021125586A1/en

[56] Robert Hackett. 2016. *LinkedIn Lost 167 Million Account Credentials in Data Breach.* Fortune. Retrieved 2021-01-14 from https://fortune.com/2016/05/18/linkedin-data-breach-email-password/

[57] Dick Hardt. 2012. The OAuth 2.0 Authorization Framework. RFC 6749. https://doi.org/10.17487/RFC6749

[58] Seongho Hong and Heeyoul Kim. 2020. VaultPoint: A Blockchain-Based SSI Model that Complies with OAuth 2.0. *Electronics* 9, 8 (2020), 1–20. https://doi.org/10.3390/electronics9081231

[59] John Hughes and Eve Maler. 2005. *Security Assertion Markup Language (SAML) V2.0 Technical Overview*. Technical Report. OASIS. Retrieved 2021-08-03 from https://www.oasis-open.org/committees/download.php/14361/sstc-saml-tech-overview-2.0-draft-08.pdf

[60] Hyperledger. 2020. Hyperledger Indy. Retrieved 2021-01-14 from https://www.hyperledger.org/use/hyperledger-indy

[61] ID123 Inc. 2021. *ID123*. ID123 Inc. Retrieved 2021-06-30 from https://www.id123.io/

[62] David Ingram. 2017. *Facebook hits 2 billion-user mark, doubling in size since 2012.* Reuters. Retrieved 2021-01-14 from https://www.reuters.com/article/us-facebook-users-idUSKBN19I2GG

[63] Koki Inoue, Dai Suzuki, Toshihiko Kurita, and Satoshi Imai. 2020. Cooperative Task Scheduling for Personal Identity Verification in Networked Systems. In *2020 32nd International Teletraffic Congress (ITC 32)*. IEEE, IEEE, Osaka, Japan, 97–105. https://doi.org/10.1109/ITC3249928.2020.00020

[64] Institute of Electrical and Electronics Engineers. 2000. IEEE Xplore. Retrieved 2021-01-14 from https://ieeexplore.ieee.org/

[65] Jim Isaak and Mina J Hanna. 2018. User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer* 51, 8 (2018), 56–59. https://doi.org/10.1109/MC.2018.3191268

[66] ISO Central Secretary. 2019. *IT Security and Privacy - A framework for identity management - Part 1: Terminology and concepts*. Standard. International Organization for Standardization, Geneva, CH. Retrieved 2021-01-14 from https://www.iso.org/standard/77582.html

[67] Philipp Jakubeit, Albert Dercksen, and Andreas Peter. 2020. SSI-AWARE: Self-sovereign Identity Authenticated Backup with Auditing by Remote Entities. In *Information Security Theory and Practice*, Maryline Laurent and Thanassis Giannetsos (Eds.). Springer International Publishing, Cham, 202–219. https://doi.org/10.1007/978-3-030-41702-4_13

[68] Zakwan Jaroucheh and Iván Abellán Álvarez. 2021. Secretation: Toward a Decentralised Identity and Verifiable Credentials Based Scalable and Decentralised Secret Management Solution. In *2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. IEEE, IEEE, Sydney, Australia, 1–9. https://doi.org/10.1109/ICBC51069.2021.9461144

[69] Audun Jøsang and Simon Pope. 2005. User Centric Identity Management. In *AusCERT Asia Pacific information technology security conference*. APCERT Secretariat, Kyoto, Japan, 77.

[70] Simon Josefsson and Ilari Liusvaara. 2017. Edwards-Curve Digital Signature Algorithm (EdDSA). RFC 8032. https://doi.org/10.17487/RFC8032

[71] Jayana Kaneriya and Hiren Patel. 2020. A Comparative Survey on Blockchain Based Self Sovereign Identity System. In *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*. IEEE, IEEE, Coimbatore, India, 1150–1155. https://doi.org/10.1109/ICISS49785.2020.9315899

[72] Christophe Kiennert, Samia Bouzefrane, and Pascal Thoniel. 2015. 3 - Authentication Systems. In *Digital Identity Management*, Maryline Laurent and Samia Bouzefrane (Eds.). Elsevier, Amsterdam, The Netherlands, 95–135. https://doi.org/10.1016/B978-1-78548-004-1.50003-1

[73] Kyung-Hoon Kim, Seungjoo Lim, Dong-Yeop Hwang, and Ki-Hyung Kim. 2021. Analysis on the Privacy of DID Service Properties in the DID Document. In *2021 International Conference on Information Networking (ICOIN)*. IEEE, IEEE, Bangkok, Thailand, 745–748. https://doi.org/10.1109/ICOIN50884.2021.9333997

[74] Won-Bin Kim, Im-Yeong Lee, and Kang-Bin Yim. 2021. Group Delegated ID-Based Proxy Re-encryption for PHR. In *Innovative Mobile and Internet Services in Ubiquitous Computing*, Leonard Barolli, Aneta Poniszewska-Maranda, and Hyunhee Park (Eds.). Springer International Publishing, Cham, 447–456. https://doi.org/10.1007/978-3-030-50399-4_43

[75] B. Kitchenham and S Charters. 2007. *Guidelines for performing systematic literature reviews in software engineering*. Technical Report. EBSE. 65 pages.

[76] Hugo Krawczyk and Tal Rabin. 1998. Chameleon Hashing and Signatures. Cryptology ePrint Archive, Report 1998/010. Retrieved 2021-08-03 from https://ia.cr/1998/010

[77] Michael Kubach and Heiko Roßnagel. 2021. A lightweight trust management infrastructure for self-sovereign identity. In *Open Identity Summit 2021*, Heiko Roßnagel, Christian H. Schunck, and Sebastian Mödersheim (Eds.). Gesellschaft für Informatik e.V., Lyngby, Denmark, 155–166. Retrieved 2021-08-03 from https://dl.gi.de/handle/20.500.12116/36489

[78] Michael Kuperberg. 2019. Blockchain-Based Identity Management: A Survey From the Enterprise and Ecosystem Perspective. *IEEE Transactions on Engineering Management* 67, 4 (2019), 1–20. https://doi.org/10.1109/TEM.2019.2926471

[79] Jan Lauinger, Jens Ernstberger, Emanuel Regnath, Mohammad Hamad, and Sebastian Steinhorst. 2021. A-PoA: Anonymous Proof of Authorization for Decentralized Identity Management. In *2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. IEEE, IEEE, Sydney, Australia, 1–9. https://doi.org/10.1109/ICBC51069.2021.9461082

[80] Gianluca Lax and Antonia Russo. 2021. A Lightweight Scheme Exploiting Social Networks for Data Minimization According to the GDPR. *IEEE Transactions on Computational Social Systems* 8, 2 (2021), 388–397. https://doi.org/10.1109/TCSS.2020.3049009

[81] Jeonghyuk Lee, Jungyeon Hwang, Jaekyung Choi, Hyunok Oh, and Jihye Kim. 2019. SIMS: Self-Sovereign Identity Management System with Preserving Privacy in Blockchain. *IACR Cryptology ePrint Archive* 1, 2019/1241 (2019), 1–13. Retrieved 2021-01-14 from https://eprint.iacr.org/2019/1241.pdf

[82] Seungjoo Lim, Min-Hyung Rhie, DongYeop Hwang, and Ki-Hyung Kim. 2021. A Subject-Centric Credential Management Method based on the Verifiable Credentials. In *2021 International Conference on Information Networking (ICOIN)*. IEEE, IEEE, Bangkok, Thailand, 508–510. https://doi.org/10.1109/ICOIN50884.2021.9333857

[83] Shu Yun Lim, Pascal Tankam Fotsing, Abdullah Almasri, Omar Musa, Miss Laiha Mat Kiah, Tan Fong Ang, and Reza Ismail. 2018. Blockchain Technology the Identity Management and Authentication Service Disruptor: A Survey. *International Journal on Advanced Science, Engineering and Information Technology* 8, 4-2 (2018), 1735–1745. https://doi.org/10.18517/ijaseit.8.4-2.6838

[84] Gregory Linklater, Alan Herbert, Christian Smith, Barry Irwin, Alan Herbert, and Barry Irwin. 2018. Toward distributed key management for offline authentication. In *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists on - SAICSIT '18* (Port Elizabeth, South Africa) *(SAICSIT '18)*. Association for Computing Machinery, New York, NY, USA, 10–19. https://doi.org/10.1145/3278681.3278683

[85] Jianwei Liu, Adam Hodges, Lucas Clay, and John Monarch. 2020. An analysis of digital identity management systems - a two-mapping view. In *2020 2nd Conference on Blockchain Research and Applications for Innovative Networks and Services (BRAINS)*. IEEE, Paris, France, 92–96. https://doi.org/10.1109/BRAINS49436.2020.9223281

[86] Yang Liu, Debiao He, Mohammad S. Obaidat, Neeraj Kumar, Muhammad Khurram Khan, and Kim-Kwang Kwang Raymond Choo. 2020. Blockchain-based identity management systems: A review. *Journal of Network and Computer Applications* 166 (9 2020), 102731. https://doi.org/10.1016/j.jnca.2020.102731

[87] Yue Liu, Qinghua Lu, Hye Young Paik, Xiwei Xu, Shiping Chen, and Liming Zhu. 2020. Design Pattern as a Service for Blockchain-Based Self-Sovereign Identity. *IEEE Software* 37, 5 (9 2020), 30–36. https://doi.org/10.1109/MS.2020.2992783 arXiv:2005.12112.

[88] Mick Lockwood. 2021. An accessible interface layer for Self-Sovereign Identity. *Frontiers in Blockchain* 3 (2021), 63. https://doi.org/10.3389/fbloc.2020.609101

[89] Devon Loffreto. 2012. *What is "Sovereign Source Authority"?* The Moxy Tongue. Retrieved 2021-01-10 from https://www.moxytongue.com/2012/02/what-is-sovereign-source-authority.html

[90] Devon Loffreto. 2013. *Administrative Precedence.* The Moxy Tongue. Retrieved 2021-01-10 from https://www.moxytongue.com/2013/01/administrative-precedence.html

[91] Devon Loffreto. 2013. *Recalibrating Sovereignty.* The Moxy Tongue. Retrieved 2021-01-10 from https://www.moxytongue.com/2013/04/recalibrating-sovereignty.html

[92] Zoltan Andras Lux, Felix Beierle, Sebastian Zickau, and Sebastian Gondor. 2019. Full-text Search for Verifiable Credential Metadata on Distributed Ledgers. In *2019 6th International Conference on Internet of Things: Systems, Management and Security, IOTSMS 2019.* IEEE, Granada, Spain, 519–528. https://doi.org/10.1109/IOTSMS48152.2019.8939249

[93] Z. A. Lux, D. Thatmann, S. Zickau, and F. Beierle. 2020. Distributed-Ledger-based Authentication with Decentralized Identifiers and Verifiable Credentials. In *2020 2nd Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS).* IEEE, Paris, France, 71–78. https://doi.org/10.1109/BRAINS49436.2020.9223292

[94] Robert C Martin, Dirk Riehle, and Frank Buschmann. 1997. *Pattern languages of program design 3.* Addison-Wesley Longman Publishing Co., Inc., Boston, USA. 529–574 pages.

[95] Simon Meier, Benedikt Schmidt, Cas Cremers, and David Basin. 2013. The TAMARIN Prover for the Symbolic Analysis of Security Protocols. In *International Conference on Computer Aided Verification.* Springer, Springer, Saint Petersburg, Russia, 696–701. https://doi.org/10.1007/978-3-642-39799-8_48

[96] Steven Melendez and Alex Pasternack. 2019. *Here are the data brokers quietly buying and selling your personal information.* Fastcompany. Retrieved 2021-01-14 from https://www.fastcompany.com/90310803

[97] John Stuart Mill. 1859. *On liberty.* John W. Parker and Son, West Strand, London, England.

[98] Teruko Miyata, Yuzo Koga, Paul Madsen, Shin-Ichi Adachi, Yoshitsugu Tsuchiya, Yasuhisa Sakamoto, and Kenji Takahashi. 2006. A Survey on Identity Management Protocols and Standards. *IEICE TRANSACTIONS on Information and Systems* 89, 1 (2006), 112–123. https://doi.org/10.1093/ietisy/e89-d.1.112

[99] Paul Mockapetris. 1983. Domain names: Concepts and facilities. RFC 882. https://doi.org/10.17487/RFC0882

[100] Sebastian Mödersheim, Anders Schlichtkrull, Georg Wagner, Stefan More, and Lukas Alber. 2019. TPL: A Trust Policy Language. In *Trust Management XIII*, Weizhi Meng, Piotr Cofta, Christian Damsgaard Jensen, and Tyrone Grandison (Eds.). Springer International Publishing, Cham, 209–223. https://doi.org/10.1007/978-3-030-33716-2_16

[101] Richard Mortier, Hamed Haddadi, Tristan Henderson, Derek McAuley, and Jon Crowcroft. 2015. Human-Data Interaction: The Human Face of the Data-Driven Society. arXiv:1412.6159 [cs.CY]

[102] Alexander Mühle, Andreas Grüner, Tatiana Gayvoronskaya, and Christoph Meinel. 2018. A survey on essential components of a self-sovereign identity. *Computer Science Review* 30 (11 2018), 80–86. https://doi.org/10.1016/j.cosrev.2018.10.002

[103] Kaiiali Mustafa and Sezer Sakir. 2021. Computer-implemented transaction system and method. https://patents.google.com/patent/WO2021064182A1/en

[104] Nitin Naik and Paul Jenkins. 2020. Self-Sovereign Identity Specifications: Govern Your Identity Through Your Digital Wallet using Blockchain Technology. In *2020 8th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud).* IEEE, Oxford, United Kingdom, 90–95. https://doi.org/10.1109/MobileCloud48802.2020.00021

[105] Aleksandr Ometov, Sergey Bezzateev, Niko Mäkitalo, Sergey Andreev, Tommi Mikkonen, and Yevgeni Koucheryavy. 2018. Multi-Factor Authentication: A Survey. *Cryptography* 2, 1 (2018), 31. https://doi.org/10.3390/cryptography2010001

[106] Kai Petersen, Robert Feldt, Shahid Mujtaba, and Michael Mattsson. 2008. Systematic Mapping Studies in Software Engineering. In *12th International Conference on Evaluation and Assessment in Software Engineering, EASE 2008.* ACM, Bari, Italy, 1–10. https://doi.org/10.14236/ewic/ease2008.8

[107] Kai Petersen, Sairam Vakkalanka, and Ludwik Kuzniarz. 2015. Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology* 64 (8 2015), 1–18. https://doi.org/10.1016/j.infsof.2015.03.007

[108] David Recordon and Drummond Reed. 2006. OpenID 2.0: a platform for user-centric identity management. In *Proceedings of the second ACM workshop on Digital identity management.* ACM, Alexandria, USA, 11–16. https://doi.org/10.1145/1179529.1179532

[109] Drummond Reed, Manu Sporny, Dave Longley, Allen Christopher, Ryan Grant, and Markus Sabadello. 2019. *Decentralized Identifiers (DIDs).* World Wide Web Consortium (W3C). Retrieved 2021-07-16 from https://www.w3.org/TR/did-core/

[110] Paul Resnick, Richard Zeckhauser, John Swanson, and Kate Lockwood. 2006. The value of reputation on eBay: A controlled experiment. *Experimental economics* 9, 2 (2006), 79–101. https://doi.org/10.1007/s10683-006-4309-2

[111] Natsuhiko Sakimura, John Bradley, Mike Jones, Breno De Medeiros, and Chuck Mortimore. 2014. *Openid Connect Core 1.0.* Technical Report. The OpenID Foundation. Retrieved 2021-01-14 from https://openid.net/specs/openid-connect-core-1_0.html

[112] Abylay Satybaldy, Mariusz Nowostawski, and Jørgen Ellingsen. 2020. *Self-Sovereign Identity Systems.* Springer International Publishing, Cham, 447–461. https://doi.org/10.1007/978-3-030-42504-3_28

[113] Miloš Savić, Mirjana Ivanović, and Lakhmi C. Jain. 2019. *Co-authorship Networks: An Introduction.* Springer International Publishing, Cham, 179–192. https://doi.org/10.1007/978-3-319-91196-0_5

[114] Karen Scarfone and Murugiah Souppaya. 2009. *Guide to Enterprise Password Management.* Technical Report. National Institute of Standards and Technology. Retrieved 2021-01-14 from https://csrc.nist.gov/publications/detail/sp/800-118/archive/2009-04-21

[115] Martin Schanzenbach. 2020. *Towards Self-sovereign, decentralized personal data sharing and identity management.* Ph.D. Dissertation. Technical University of Munich, Germany. Retrieved 2021-08-03 from https://mediatum.ub.tum.de/1545514

[116] Martin Schanzenbach, Georg Bramm, and Julian Schütte. 2018. reclaimID: Secure, Self-Sovereign Identities Using Name Systems and Attribute-Based Encryption. In *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*. IEEE, New York, USA, 946–957. https://doi.org/10.1109/TrustCom/BigDataSE.2018.00134

[117] Martin Schanzenbach., Thomas Kilian., Julian Schütte., and Christian Banse. 2019. ZKlaims: Privacy-preserving Attribute-based Credentials using Non-interactive Zero-knowledge Techniques. In *Proceedings of the 16th International Joint Conference on e-Business and Telecommunications - SECRYPT,*. INSTICC, SciTePress, Prague, Czech Republic, 325–332. https://doi.org/10.5220/0007772903250332

[118] Frederico Schardong and Ricardo Custódio. 2021. Study Selection Process, Spreadsheet. https://docs.google.com/spreadsheets/d/1FzUJRqe3WUhtsNYV6PyMZO8F14iQO-DX

[119] C. P. Schnorr. 1990. Efficient Identification and Signatures for Smart Cards. In *Advances in Cryptology — CRYPTO' 89 Proceedings*, Gilles Brassard (Ed.). Springer New York, New York, NY, 239–252. https://doi.org/10.1007/0-387-34805-0_22

[120] Matthew Schutte. 2016. Schutte's Critique of the Self-Sovereign Identity Principles. Retrieved 2021-01-14 from http://matthewschutte.com/2016/10/25/schuttes-critique-of-the-self-sovereign-identity-principles/

[121] David Searls. 2012. *The Identity Problem*. Project VRM. Retrieved 2021-01-10 from https://blogs.harvard.edu/vrm/2012/11/08/the-identity-problem/

[122] David Searls. 2013. *IIW Challenge #1: Sovereign Identity in the Great Silo Forest*. Doc Searls Weblog. Retrieved 2021-01-10 from http://blogs.harvard.edu/doc/2013/10/14/iiw-challenge-1-sovereign-identity-in-the-great-silo-forest/

[123] David Searls. 2013. *Leveraging Whitman*. Project VRM. Retrieved 2021-01-10 from http://blogs.harvard.edu/vrm/2013/08/21/leveraging-whitman/

[124] David Searls. 2014. *Personal = Sovereign*. Project VRM. Retrieved 2021-01-10 from http://blogs.harvard.edu/vrm/2014/02/06/personal-sovereign/

[125] David Searls. 2018. *Some Perspective on Self-Sovereign Identity*. Kuppingercole. Retrieved 2021-01-10 from https://www.kuppingercole.com/blog/guest/some-perspective-on-self-sovereign-identity

[126] Adi Shamir. 1979. How to share a secret. *Commun. ACM* 22, 11 (1979), 612–613. https://doi.org/10.1145/359168.359176

[127] Yashothara Shanmugarasa, Hye-Young Paik, Salil S. Kanhere, and Liming Zhu. 2021. Towards Automated Data Sharing in Personal Data Stores. In *2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. IEEE, IEEE, Pisa, Italy, 328–331. https://doi.org/10.1109/PerComWorkshops51409.2021.9431001

[128] Har Preet Singh, Kyriakos Stefanidis, and Fabian Kirstein. 2021. A Private Key Recovery Scheme Using Partial Knowledge. In *2021 11th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*. IEEE, IEEE, Paris, France, 1–5. https://doi.org/10.1109/NTMS49979.2021.9432642

[129] Samuel M. Smith. 2019. Key Event Receipt Infrastructure (KERI). *CoRR* abs/1907.02143 (2019), 140. arXiv:1907.02143 Retrieved 2021-08-03 from http://arxiv.org/abs/1907.02143

[130] Reza Soltani, Uyen Trang Nguyen, and Aijun An. 2019. Practical Key Recovery Model for Self-Sovereign Identity Based Digital Wallets. In *Proceedings - IEEE 17th International Conference on Dependable, Autonomic and Secure Computing, IEEE 17th International Conference on Pervasive Intelligence and Computing, IEEE 5th International Conference on Cloud and Big Data Computing, 4th Cyber Scienc*. IEEE, Fukuoka, Japan, 320–325. https://doi.org/10.1109/DASC/PiCom/CBDCom/CyberSciTech.2019.00066

[131] Tim Speelman. 2020. *Self-Sovereign Identity: Proving Power over Legal Entities*. Master's thesis. Delft University of Technology. Retrieved 2021-01-14 from https://repository.tudelft.nl/islandora/object/uuid:aab1f3ff-da54-47f7-8998-847cb78322c8

[132] Manu Sporny, Dave Longley, and David Chadwick. 2017. *Verifiable Credentials Data Model 1.0*. World Wide Web Consortium (W3C). Retrieved 2021-01-14 from https://www.w3.org/TR/vc-data-model/

[133] Manu Sporny, Dave Longley, and David Chadwick. 2019. *Verifiable Credentials Data Model 1.0 - Trust Model*. World Wide Web Consortium (W3C). Retrieved 2021-01-14 from https://www.w3.org/TR/vc-data-model/#trust-model

[134] Manu Sporny, Dmitri Zagidulin, and Dave Longley. 2019. *The did:key Method*. World Wide Web Consortium (W3C). Retrieved 2021-08-04 from https://w3c-ccg.github.io/did-method-key/

[135] Springer Nature. 2012. Springer Link. Retrieved 2021-01-14 from https://link.springer.com/

[136] William Stallings. 2013. *Cryptography and Network Security: Principles and Practice* (6th ed.). Prentice Hall Press, USA.

[137] Quinten Stokkink and Johan Pouwelse. 2018. Deployment of a Blockchain-Based Self-Sovereign Identity. In *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. IEEE, Halifax, NS, Canada, 1336–1342. https://doi.org/10.1109/Cybermatics_2018.2018.00230

[138] Melanie Swan. 2015. *Blockchain: Blueprint for a new economy*. "O'Reilly Media, Inc.", Sebastopol, USA.

[139] Shuhaili Talib, Nathan L Clarke, and Steven M Furnell. 2010. An Analysis of Information Security Awareness within Home and Work Environments. In *2010 International Conference on Availability, Reliability and Security*. IEEE, IEEE, Krakow, Poland, 196–203. https://doi.org/10.1109/ARES.2010.27

[140] The Sovrin Foundation. 2016. Sovrin. Retrieved 2021-01-14 from https://sovrin.org/

[141] The Sovrin Foundation. 2016. Write To The Sovrin Public Ledger. Retrieved 2021-01-14 from https://sovrin.org/issue-credentials/

[142] Andrew Tobin and Drummond Reed. 2016. *The Inevitable Rise of Self-Sovereign Identity*. Technical Report. The Sovrin Foundation. Retrieved 2021-01-14 from https://sovrin.org/wp-content/uploads/2018/03/The-Inevitable-Rise-of-Self-Sovereign-Identity.pdf

[143] Kalman C. Toth and Alan Anderson-Priddy. 2019. Self-Sovereign Digital Identity: A Paradigm Shift for Identity. *IEEE Security and Privacy* 17, 3 (5 2019), 17–27. https://doi.org/10.1109/MSEC.2018.2888782

[144] Kalman C Toth, Ann Cavoukian, and Alan Anderson-Priddy. 2020. Privacy by Design Identity Architecture Using Agents and Digital Identities. In *Annual Privacy Forum*. Springer, Springer, Lisbon, Portugal, 73–94. https://doi.org/10.1007/978-3-030-55196-4_5

[145] Sean Turner, Stephen Farrell, and Russ Housley. 2010. An Internet Attribute Certificate Profile for Authorization. RFC 5755. https://doi.org/10.17487/RFC5755

[146] uPort. 2020. uPort. Retrieved 2021-01-14 from https://www.uport.me/

[147] William Uzgalis. 2020. John Locke. In *The Stanford Encyclopedia of Philosophy* (Spring 2020 ed.), Edward N. Zalta (Ed.). Metaphysics Research Lab, Stanford University, Stanford, USA.

[148] Dirk van Bokkem, Rico Hageman, Gijs Koning, Luat Nguyen, and Naqib Zarin. 2019. Self-Sovereign Identity Solutions: The Necessity of Blockchain Technology. *CoRR* abs/1904.12816 (2019), 1–8. arXiv:1904.12816 http://arxiv.org/abs/1904.12816

[149] Marvin Van Wingerde. 2017. *Blockchain-enabled Self-sovereign Identity*. Master's thesis. Tilburg University. https://doi.org/10.13140/RG.2.2.17693.82406

[150] Mountain View, Johan Rydell, Mingliang Pei, and Salah Machani. 2011. TOTP: Time-Based One-Time Password Algorithm. RFC 6238. https://doi.org/10.17487/RFC6238

[151] W3C Technology and society domain. 2017. *Verifiable Claims Working Group Frequently Asked Questions*. World Wide Web Consortium (W3C). Retrieved 2021-01-14 from https://w3c.github.io/webpayments-ig/VCTF/charter/faq.html#self-sovereign

[152] Matthias Wachs, Martin Schanzenbach, and Christian Grothoff. 2013. On the feasibility of a censorship resistant decentralized name system. In *International Symposium on Foundations and Practice of Security*. Springer, Springer, La Rochelle, France, 19–30. https://doi.org/10.1007/978-3-319-05302-8_2

[153] Brent Waters. 2005. Efficient identity-based encryption without random oracles. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, Springer, Aarhus, Denmark, 114–127. https://doi.org/10.1007/11426639_7

[154] Ryan West, Daniel Bluhm, Matthew Hailstone, Stephen Curran, Sam Curren, and George Aristy. 2019. *Aries RFC 0023: DID Exchange Protocol 1.0*. Technical Report. Linux Foundation. Retrieved 2021-07-16 from https://github.com/hyperledger/aries-rfcs/blob/master/features/0023-did-exchange/README.md

[155] Ryan West, Daniel Bluhm, Matthew Hailstone, Stephen Curran, Sam Curren, and George Aristy. 2019. *Aries RFC 0434: Out-of-Band Protocol 1.1*. Technical Report. Linux Foundation. Retrieved 2021-08-04 from https://github.com/hyperledger/aries-rfcs/blob/master/features/0434-outofband/README.md

[156] Sven Wohlgemuth, Katsuyuki Umezawa, Yusuke Mishina, and Kazuo Takaragi. 2020. A Secure Decision-Support Scheme for Self-Sovereign Identity Management. In *Symposium on Cryptography and Information Security (SCIS)*. IEICE, Kochi, Japan, 1–8.

[157] Claes Wohlin. 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering - EASE '14*. ACM Press, London, England, 1–10. https://doi.org/10.1145/2601248.2601268

[158] Jie Xu, Kaiping Xue, Hangyu Tian, Jianan Hong, David S.L. Wei, and Peilin Hong. 2020. An Identity Management and Authentication Scheme Based on Redactable Blockchain for Mobile Networks. *IEEE Transactions on Vehicular Technology* 69, 6 (6 2020), 6688–6698. https://doi.org/10.1109/TVT.2020.2986041

[159] Kaliya Young and Infominer. 2021. *Self-Sovereign Identity: Decentralized digital identity and verifiable credentials*. Vol. 1. Manning, Shelter Island, ÜSA, Chapter The origins of the SSI community, 310–321. Retrieved 2021-01-10 from https://livebook.manning.com/book/self-sovereign-identity/chapter-16/

[160] Zibin Zheng, Shaoan Xie, Hong-Ning Dai, Weili Chen, Xiangping Chen, Jian Weng, and Muhammad Imran. 2020. An overview on smart contracts: Challenges, advances and platforms. *Future Generation Computer Systems* 105 (2020), 475–491. https://doi.org/10.1016/j.future.2019.12.019

[161] Xiaoyang Zhu and Youakim Badr. 2018. Fog computing security architecture for the internet of things using blockchain-based social networks. In *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. IEEE, IEEE, Halifax, NS, Canada, 1361–1366. https://doi.org/10.1109/Cybermatics_2018.2018.00234

[162] Xiaoyang Zhu and Youakim Badr. 2018. Identity Management Systems for the Internet of Things: A Survey Towards Blockchain Solutions. *Sensors* 18, 12 (2018), 4215. https://doi.org/10.3390/s18124215