Prison Break: From Proprietary Data Sources To SSI Verifiable Credentials

Katja Assaf, Alexander Mühle, Daniel Köhler, and Christoph Meinel

Abstract Despite extensive efforts, smaller companies and organisations often fail to be GDPR compliant. GDPR demands that the data subject's information is available to the data subject in a simple and structured way. One option to provide the data with additional benefits is issuing verifiable credentials (VCs) following the W3C standard and, thus, introducing the data provider as an issuer into a Self-Sovereign Identity (SSI) system. We show that this can be achieved with limited overhead by introducing a middleware component, which is only loosely coupled with the existing ecosystem. To enhance user acceptance, we define our design goals as usability, security, and privacy, which we manage to achieve partially. During our work, we identified several challenges, such as revocation, verifiability of verifiers, and legal regulations, which provide options for future research in developing Self-Sovereign Identity solutions towards real-world applicability.

Katja Assaf · Alexander Mühle · Daniel Köhler · Christoph Meinel
Hasso Plattner Institute
Potsdam, Germany
e-mail: katja.assaf@hpi.de

This version of the contribution has been accepted for publication, after peer review but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/10.1007/978-3-031-28451-9_31. Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms

1 Introduction

Since May 2018, the General Data Protection Regulation (GDPR) has been binding for all European Union member states. It requires businesses with customers within the European Union to process data in a secure, transparent and accountable way. However, according to [7], roughly half of the asked small businesses are not yet GDPR compliant despite investing heavily in compliance. In contrast, big companies such as google already provide specialised tools to enable customers to create a (data) takeout, which is a downloadable package of the user's data. Two central points within GDPR are Article 15 *Right of access by the data subject* and Article 20 *Right to data portability*. The *right of access by the data subject* requires organisations to make user data available to the data subject after authentication. The *right to data portability* goes further and requires organisations to provide the data in a "structured, commonly used and machine-readable format" [5]. We suggest enhancing GDPR compliance by providing a data takeout in a verifiable format, which grants data subjects the additional benefit of being reusable within the Self-Sovereign Identity (SSI) system. Verifiable credentials (VCs) are beneficial as learner records, health data, e.g. vaccination certificates or customer data, although they require more caution from a data privacy perspective.

Self-Sovereign Identity (SSI) is a design principle that emphasises the user's control over their data. It is at the stage where a minimum user base still needs to be established [9]. In addition to establishing a standard, as currently done by the W3C[1], it is crucial to enable individuals and organisations to participate in the system with minimal friction. Besides raising the number of credential holders and verifiers, also called relying parties, the number of credential issuers within the system also needs to rise. According to Schmidt et al. [19], in 2021, only 11 out of 147 investigated projects were classified as issuers. Consequently, we will answer the question:

> **RQ:** How to enable a service provider to become an issuer in an SSI system and, thus, integrate an SSI system with an existing ecosystem with minimal friction?

Organisations already have an infrastructure for identity and customer information management, and these existing data sources are the logical starting point for a potential credential issuance process. For this purpose, we present an approach utilising existing identity and customer information management systems (as shown in figure 1) for issuing VCs, lowering the barrier of entry for issuers to SSI ecosystems.

Self-Sovereign Identity (SSI), as stated above, is a design goal for an identity management system giving the credential holder full control over their digital identity [13]. However, SSI often refers to concrete implementations that usually rely on blockchain and only partially fulfil the design goals at best. These self-claimed SSI systems differ from what we mean when referring to an SSI system.
An SSI system consists of three actors: *issuer*, *relying party* and *credential holder*. The credential holder owns credentials containing claims about them or entities related to them. The credential holder receives or retrieves credentials from an issuer,

[1] https://www.w3.org/TR/vc-data-model/

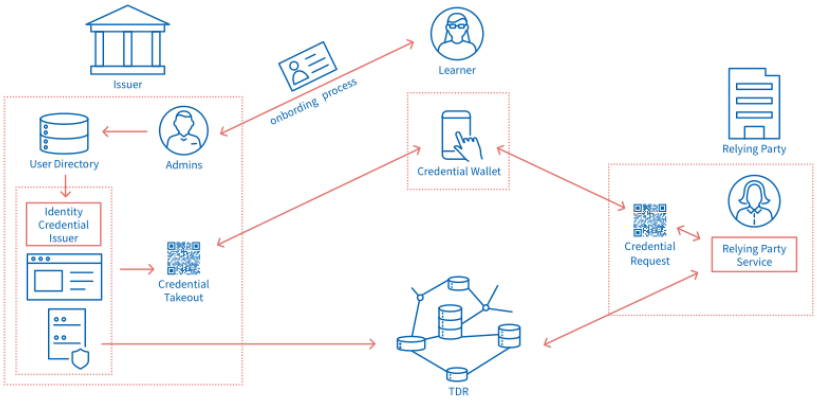

Fig. 1 System Overview

who signs the claim. The holder bakes one or more credentials into a credential presentation and presents the presentation to a relying party, which verifies all signatures and looks up the current status of the entities in a trusted data registry.

Due to the primary purpose of the relying party, it is called *a verifier* in particular contexts. We will use both terms interchangeably to emphasise either the role within the system or the protocol carried out. We will use the term *customer* for individuals who have an established relationship with an organisation but have yet to become actors within the SSI system. Additionally, we will use the term *user* to refer to any participant of an SSI system regardless of its specific role.

Customer Information Management is necessary for all services, especially those available online, allowing customers to access their data via a login-protected website. This can be access to a bank account, the last eCommerce order or learning achievements at a learning platform. These service providers have in common that they store data in their database and run a web server connected to their database and an identity provider, which checks access control. Most use the same communication standards: oAuth 2.0 is the standard for federated identity management, meaning it is used for signing in with Facebook, Google or Twitter on another web page. Access to the data is provided via a RESTful API, although many are not publicly available but restricted to the company's network.

Our Contribution To enable organisations to offer SSI credentials to their customers and enhance their GDPR compliance, we propose a new scheme to enable users and organisations to transform existing data into SSI credentials. As described above, identity management systems already have a harmonised landscape for accessing user profiles. In our proof of concept, we show the viability of using OIDC to enable organisations as credential issuers easily. We restrict assumptions about the SSI system and the organisation's infrastructure to a minimum (section 3.2) to offer a solution as generic as possible in compliance with our defined design goals

(section 3.1). We describe our reference implementation to ensure the feasibility of our suggestion (section 3.3) and assess our solution against our design goals while pointing out its limitations (section 4).

2 Related Work

In general, research interest in Self-Sovereign Identity steadily increased in recent years from 5 papers in 2018 to 37 in 2021, according to Schardong et al. [18]. Additionally, Schardong et al. [18] provide a taxonomy to classify practical problems discussed in the literature. The work of Grüner et al. [9] is classified as protocol integration, identity derivation and trust policy evaluation. Grüner et al. [9] tackle the problem of achieving a minimal user base in Self-Sovereign Identity. Thus, having a similar focus as our work but considering it from a different angle. While our solution strives to include an organisation with existing data as an issuer, Grüner et al. enable an organisation as a verifier. In terms of classical identity management, we can say that Grüner et al. developed a solution for federated identity verification, while our solution describes federated identity issuance.

Regarding the taxonomy developed by Schardong et al. [18], our solution would also fit into the category of protocol integration. However, from the eight papers in this category, only Jurado et al. [12] consider integrating an existing data provider as an issuer into the SSI ecosystem, while the other seven paper work on enabling service providers as SSI verifiers by enabling authentication with SSI. Jurado et al. [12] are firmly set in the European Health Insurance Card (EHIC) use case and its integration with eIDAS. Thus, their solution contains more details than ours and considers issuer and verifier integration simultaneously, as well as the underlying trust framework. In contrast, our solution makes fewer assumptions on the existing infrastructure and tries to provide a more flexible model focused on only issuer integration.

The survey of Schardong et al. [18] covers nearly all of the relevant literature since it was published in August 2022. The existence of two more research papers [11], [1] advancing the research on SSI integration, also published last summer and thus not considered within the survey, emphasises computer scientists' current interest in the topic.

Kuperberg et al. [11] provide a state-of-the-art survey about bridging the gap between SSI and traditional IAM systems. Most of the over 40 SSI solutions were excluded since they were not concerned with integration. The remaining sources were either concerned with authentication via SSI by enabling the identity provider (Pattern A) or translating VCs into the OIDC protocol (Pattern B and C).

Bolgouras et al. integrated the FIDO protocol and the eIDAS framework into an SSI framework to allow for authentication using FIDO and verification across countries [1], extending authentication options for SSI solutions.

3 Proof of Concept

To show the feasibility of our idea, we developed a proof of concept enabling a MOOC provider to issue verifiable credentials (VCs). Since we strive to achieve flexibility and broaden user acceptance, we defined our design goals accordingly and restricted assumptions to the necessary minimum.

3.1 Design Goals

Usability Our overall goal is to provide a solution allowing users to retrieve their existing data as a VC. Since we aim for a wider adoption of SSI and most users are reluctant to change, usability is deemed crucial. While this topic is typically viewed from the credential holder's perspective, we focus on the issuer instead. While the holder requesting the VC needs an intuitive way of performing the request, storing received credentials and managing them securely, this is a topic for a separate research project. As we aim to increase the number of active issuers, the administrators in charge of the organisation's network require a service which is easy to integrate and maintain.

DG 1: Integration with an existing network shall be frictionless.

Privacy and Security Using SSI as an individual (credential holder) is a means to gain more control over one's data and thus protect one's privacy. Otherwise, federated identity management with single sign-on would be a better solution since it is a mature system providing good usability in general. With SSI, the responsibility for the security and privacy of the data is shifted from a trusted third party to the individual user, the credential holder. Thus, it is necessary to design a system which guides the user and minimises the options for bad choices, such as publishing private keys.

DG 2: The proof of concept architecture shall follow the security-by-design principle.
DG 3: To ensure the holder's privacy, they shall have full control over their data.

3.2 Assumptions and Design Decisions

For the proof of concept, we decided to utilise a micro-service architecture, keeping the integration efforts minimal and allowing the replacement of components with any potentially already existing systems of the implementing organisation later on [10]. Modularisation also allows us to build connectors more easily with the organisation's existing infrastructure. The key goal of the proof of concept is to develop approaches to integrate existing systems and processes. For this purpose, we took the following assumptions about the organisations:

AS 1: Customers (potential credential holders) are already known to the organisation and are managed in a user database.
AS 2: Authentication is realised via OpenIDConnect[2] (OIDC).
AS 3: Customer data is accessible through a well-defined API.

3.3 Architecture

We developed a service that enables customers to receive two types of credentials. A high level overview can be seen in figure 2. On the one hand, the new credential holder can receive so-called identity credentials, encapsulating their base user profile data such as name, birth date or other personally identifiable information (PII). Additionally, we integrated an example of customer data accessible through an API, in our case learning achievement data of openHPI[3].

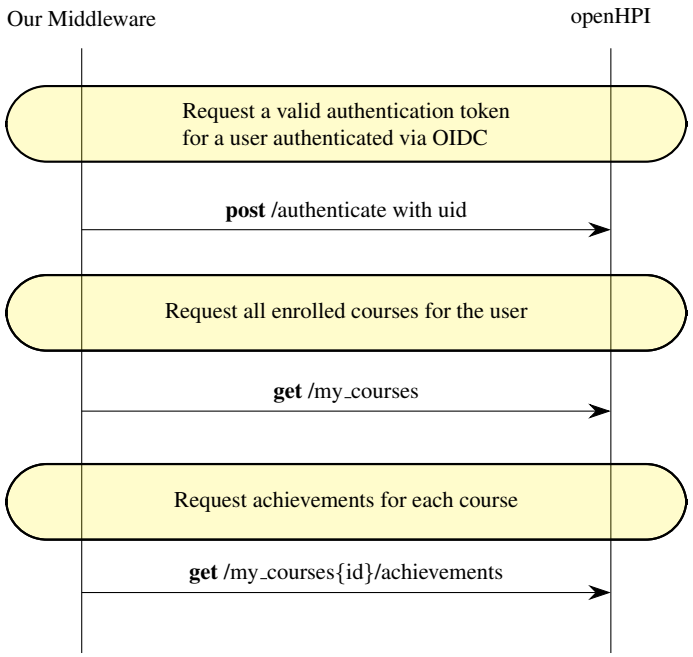

Fig. 2 Retrieving Customer Data From openHPI

The basic functionalities needed for issuing and verifying VCs of arbitrary content are core services in every such deployment. Therefore, the interfaces of these

[2] OpenIDConnect is a widely used extension of the oAuth 2.0 protocol.
[3] open.hpi.de

services are already standardised by the VC-API working group[4]. We have implemented these basic building blocks compliant with the standard. As we have utilised a micro-service architecture, we have a separate service (issuer middleware) for interacting with the credential holder and handling customer data from other services. The issuer middleware will use two data sources for our proof of concept: an identity provider and the openHPI system.

Identity Provider Identity management protocols have converged towards a few established ones. Especially on the web, OIDC has emerged as the prevailing solution for federated identity management. However, these identity providers cannot issue VCs necessary for customers to participate in the SSI ecosystem. In our proof of concept, we utilise Keycloak[5] as an example of a popular identity provider that supports OIDC, SAML and other identity management protocols. The user database of Keycloak was filled with test data, mimicking the user management of an organisation. However, in a real-world deployment, this process would be defined by the organisation's onboarding process. Long-standing onboarding procedures might require in-person contact to fulfil specific bureaucratic needs and can be implemented upstream. In the end, user data will be put into a database and integrated with an identity provider.

openHPI API While interacting with identity management data has been relatively homogenised through standard protocols such as OIDC, customer data is much more diverse, and the same kind of overarching protocol rarely exists. Each use case and organisation has its own APIs exposing user-relevant data. As an example of integrating this kind of data source, we have integrated openHPI, an open online learning platform. From this platform, we can retrieve information on enrolled courses, details of these courses and, most importantly, for the issuance of verifiable credentials, the learning achievements earned by the learner. The learner, in our setting, is the customer and, as such, the credential holder of the SSI system.

Initially, we utilise the user ID of the learner, which we have previously obtained through single sign-on with the identity provider, to receive a valid authentication token. The token can be used to request all enrolled courses for the learner. The course data is then enriched with further details from the general course API so that a visual choice of available courses can be presented to the learner. With a further request, the achievement for each course can be retrieved, which is then available for the learner to select.

Credential Holder Flow The process of receiving a VC from an established data source is described in figure 3. When requesting an identity credential or a VC from the data sources, the credential holder visits the frontend of our middleware. Here they need to authenticate themselves using the Keycloak identity provider. Once they have successfully done so, their profile is available to the middleware. The credential holder is presented with the possible data to export using the retrieved profile. After selecting the desired attributes, the VC issuance process itself is started. Once

[4] w3c-ccg.github.io/vc-api/

[5] keycloak.org

the middleware has issued the VC using the Issuer Service, it is sent to the credential holder for storing in a wallet.

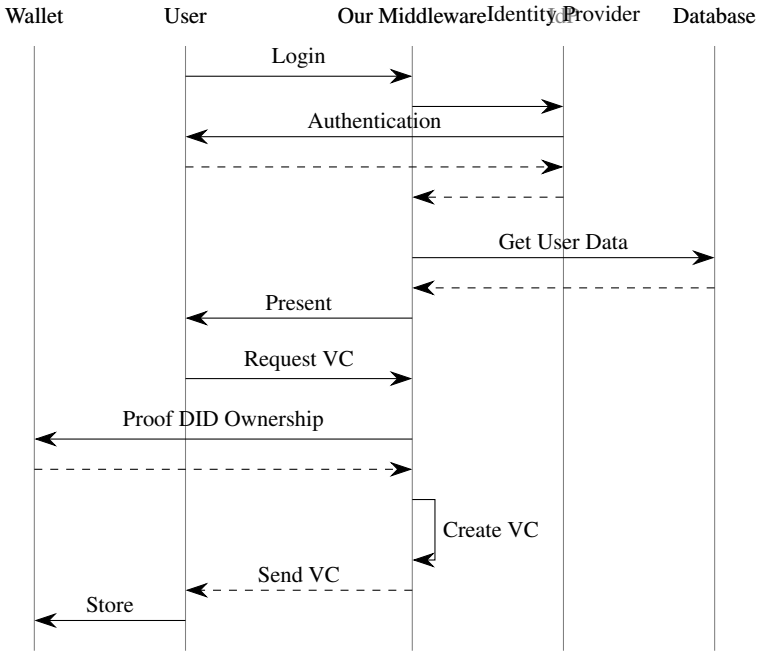

Fig. 3 Sequence Diagram: Data Flow

4 Discussion and Improvements

As discussed below, we achieved our design goals as far as practicable in our restricted proof of concept setting. Although our work is a reasonable first step towards federated identity issuance in SSI, we identified additional technical and conceptual challenges, providing options for further research.

4.1 Implementation of Design Goals

Usability Following the argumentation of Venters et al. [21], usability is a requirement for sustainable software. In the sense of Venters et al. [20], sustainable software is software capable of being maintained and, thus, continues to exist. From a software architecture standpoint, *separation of concerns* is required ([21], [6]), an

idea describing the modularisation of software. Software is broken into modules according to their functionality and with well-defined interfaces between the modules.

Consequently, the focus on modularisation of our solution by introducing an additional middleware improves sustainability and, thus, usability. The existing organisation's infrastructure does not need an extension but rather an additional configuration. Fewer changes to the software require less testing, documentation, and verification during the implementation and set-up phase, limiting cost and effort for a potential new issuer compared to a system extension.

Security Our solution's security heavily depends on the organisation's network security. Optimally, the API towards the database allows exactly one necessary request, which fetches the data of the currently logged-in customer. Further, our solution uses only the standard functionality of the identity provider. Thus, the damage that can be done to the organisation's network if our solution is corrupted is limited to stealing the data of customers whose key material has been compromised.

A bigger problem is the access to the organisation's signing key. If our solution gets compromised, so is the signing key. Consequently, revocation of all certificates issued after the attack is crucial.

Privacy As shown in figure 3, no additional data is added directly to the organisation. However, the organisation can collect metadata about the customer with every request sent, but the same holds true if the customer requests data via the organisation's website. Thus, we consider the customer's privacy risk towards the organisation a minor one for our use case.

Additionally, the VC holds a risk for the holder's privacy since the available data is now verifiable, making it more valuable to data collectors and thieves. If the data is not verifiable, many customers give false information to protect their privacy [16]. Again this holds true for all certifications in general, and thus SSI systems in particular. Credential holders need to be educated on the value of data and how to protect it. As a first step, our system provides a technical solution for enabling the individual toward data protection. In traditional systems, the content of a credential is entirely under the issuer's control. In contrast to the traditional system, the credential holder chooses which data to combine into a VC in our solution. Thus, they can leave out critical data they do not want to disclose from the beginning or request different VCs for different purposes.

4.2 Known Limitations

Our proposed solution is a proof of concept and will need adaptions before it is deployable for real-world applications. Some of the identified shortcomings, such as *revocation* and *credential management*, are technical and can be solved on an individual project level. Others, such as *semantics standardisation* and *legal regulations*, require a broader approach pushed forward by official institutions. *Verifiability of verifiers* can be seen as an in-between since it is possible to solve for an individ-

ual project. At the same time, an overarching approach would be more efficient and favourable in terms of interoperability.

Revocation As seen in section 4.1, revocation is a necessary feature to mitigate the threat of a compromised signing key. Revocation is still an open research field in cryptography ([22], [8], [4]) as well as in SSI ([14], [3]), with many possible solutions known. They all have advantages and drawbacks, often regarding performance versus privacy concerns. Thus, a general recommendation cannot be given.

Credential Management As control is shifted from the issuer to the credential holder, so is responsibility. A credential holder has to manage key material and backups for secure and privacy-preserving usage of an SSI system. Therefore wallets are employed. Despite their increased popularity in recent years, wallets are still in an early stage and need to improve, especially in usability and interoperability [17]. At the time of writing, the W3C Universal Wallet specification is still in a draft state[6]. One of the more mature solutions is the DCC's Learner Credential Wallet[7]. However, the Learner Credential Wallet focuses on the academic credential use case, and consequently, it is not sufficient for all use cases.

Verifiable Verifiers Following the principles of SSI, it is favourable that verifiers or relying parties requesting data from a credential holder authenticate themselves to the credential holder. The relying party's authentication would empower credential holders to make informed decisions about with whom they share which data. Consequently, a complete infrastructure to manage public keys, establish trust in relying parties and supervise, which attributes a relying party can request, would be required.

Semantics Standardisation The claims in a VC are stored in a JSON[8] or JSON-LD[9] file. The content of a JSON can be automatically validated with the help of schemas[10], while JSON-LD offers less flexibility but semantic interoperability. Different fields are working on schemas fitting their needs, such as the ELMO-xml format in the Education sector. However, most standardisation activities are still in a draft state.

Legal Regulations The usefulness of digital certificates is foremost dependent on laws and regulations, despite the topic not being widely recognised in research. However, Brown [2] described the possible advantages of digital signatures over handwritten signatures in 1993. According to [2], digital signatures provide a higher level of assurance and can supplement or even surpass the analogue form if they get legally recognised. Pattiyanon and Aoki [15] performed a systematic review of laws, regulations and standards applicable to SSI in general, excluding domain-specific regulations. Within the 28 sources identified, 8 were only applicable on a national

[6] https://w3id.org/wallet

[7] https://github.com/digitalcredentials/learner-credential-wallet

[8] https://www.json.org/

[9] https://json-ld.org/

[10] https://w3c-ccg.github.io/vc-json-schemas/

level, fragmenting the legal landscape further. The eIDAS regulation has tried to lay down a common legal ground for digital signatures within the EU over the past years. However, the acceptance of digital signatures is only increasing slowly. Still, the increase is an accomplishment, improving the situation from a fragmented international regulation landscape in Europe towards a more homogeneous acceptance.

5 Conclusion

We presented a proof of concept for integrating an organisation into the SSI ecosystem as a credential issuer, which can be a building block for achieving a minimum user base for SSI systems. We tested our solution by integrating it with a MOOC platform.

However, additional real-world integration projects are necessary to solve the technical known limitations 4.2 and identify further challenges when integrating existing infrastructure with SSI. The presented proof of concept can encourage organisations to put SSI to the test with limited risk since our solution is only loosely coupled with the existing organisations' infrastructure and evaluate whether being an SSI credential issuer provides additional value for them and their customers.

Acknowledgements This work has been funded through the Federal Ministry for Education and Research (BMBF) under grant M534800. We want to thank our partners at the TU Munich and the German Academic Exchange Service (DAAD) for the discussions on the topic.

References

1. Bolgouras, V., Angelogianni, A., Politis, I., Xenakis, C.: Trusted and secure self-sovereign identity framework. In: Proceedings of the 17th International Conference on Availability, Reliability and Security. pp. 1–6 (2022)
2. Brown, P.W.: Digital signatures: can they be accepted as legal signatures in edi? In: Proceedings of the 1st ACM conference on Computer and communications security. pp. 86–92 (1993)
3. Chotkan, R., Decouchant, J., Pouwelse, J.: Distributed attestation revocation in self-sovereign identity. In: 2022 IEEE 47th Conference on Local Computer Networks (LCN). pp. 414–421. IEEE (2022)
4. Emura, K., Takayasu, A., Watanabe, Y.: Generic constructions of revocable hierarchical identity-based encryption. Cryptology ePrint Archive (2021)
5. EU: Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation). Tech. rep., European Union (2016)
6. Garlan, D.: Software architecture: a roadmap. In: Proceedings of the Conference on the Future of Software Engineering. pp. 91–101 (2000)
7. gdpr.eu: 2019 gdpr small business survey. Tech. rep., Proton AG (2019), `https://gdpr.eu/wp-content/uploads/2019/05/2019-GDPR.EU-Small-Business-Survey.pdf`
8. Ge, A., Wei, P.: Identity-based broadcast encryption with efficient revocation. In: IACR International Workshop on Public Key Cryptography. pp. 405–435. Springer (2019)

9. Grüner, A., Mühle, A., Meinel, C.: An integration architecture to enable service providers for self-sovereign identity. In: 2019 IEEE 18th International Symposium on Network Computing and Applications (NCA). pp. 1–5. IEEE (2019)
10. Jamshidi, P., Pahl, C., Mendonça, N.C., Lewis, J., Tilkov, S.: Microservices: The journey so far and challenges ahead. IEEE Software **35**(3), 24–35 (2018)
11. Kuperberg, M., Klemens, R.: Integration of self-sovereign identity into conventional software using established iam protocols: A survey. Open Identity Summit 2022 (2022)
12. Martinez Jurado, V., Vila, X., Kubach, M., Henderson Johnson Jeyakumar, I., Solana, A., Marangoni, M.: Applying assurance levels when issuing and verifying credentials using trust frameworks. Open Identity Summit 2021 (2021)
13. Mühle, A., Grüner, A., Gayvoronskaya, T., Meinel, C.: A survey on essential components of a self-sovereign identity. Computer Science Review **30**, 80–86 (2018)
14. Mühle, A., Hoops, F., Assaf, K., Meinel, C.: Manuscript: Universal statuslist: Making a case for more middleware in self-sovereign identity (2023)
15. Pattiyanon, C., Aoki, T.: Compliance ssi system property set to laws, regulations, and technical standards. IEEE Access (2022)
16. Polat, H., Du, W.: Svd-based collaborative filtering with privacy. In: Proceedings of the 2005 ACM symposium on Applied computing. pp. 791–795 (2005)
17. Sartor, S., Sedlmeir, J., Rieger, A., Roth, T.: Love at first sight? a user experience study of self-sovereign identity wallets. In: ECIS 2022 Proceedings (2022)
18. Schardong, F., Custódio, R.: Self-sovereign identity: A systematic review, mapping and taxonomy. Sensors **22**(15), 5641 (2022)
19. Schmidt, K., Mühle, A., Grüner, A., Meinel, C.: Clear the fog: Towards a taxonomy of self-sovereign identity ecosystem members. In: 2021 18th International Conference on Privacy, Security and Trust (PST). pp. 1–7. IEEE (2021)
20. Venters, C., Lau, L., Griffiths, M., Holmes, V., Ward, R., Jay, C., Dibsdale, C., Xu, J.: The blind men and the elephant: Towards an empirical evaluation framework for software sustainability. Journal of Open Research Software **2**(1) (2014)
21. Venters, C.C., Capilla, R., Betz, S., Penzenstadler, B., Crick, T., Crouch, S., Nakagawa, E.Y., Becker, C., Carrillo, C.: Software sustainability: Research and practice from a software architecture viewpoint. Journal of Systems and Software **138**, 174–188 (2018)
22. Yu, T., Xie, H., Liu, S., Ma, X., Jia, X., Zhang, L.: Certrevoke: a certificate revocation framework for named data networking. In: Proceedings of the 9th ACM Conference on Information-Centric Networking. pp. 80–90 (2022)