Information & Management 59 (2022) 103553

Contents lists available at ScienceDirect

Information & Management

journal homepage: www.elsevier.com/locate/im

Designing a Framework for Digital KYC Processes Built on Blockchain-Based Self-Sovereign Identity

Vincent Schlatt [a,b,c,*], Johannes Sedlmeir [a,b,c], Simon Feulner [a,b,c,d], Nils Urbach [a,c,d]

[a] Project Group Business & Information Systems Engineering of the Fraun-hofer FIT, Germany
[b] University of Bayreuth, Germany
[c] Research Center Finance & Information Management, Germany
[d] Frankfurt University of Applied Sciences, Germany

ARTICLE INFO

Keywords:
Banking
Digital certificate
Digital wallet
Decentralized identity
Distributed ledger technology
Verifiable credential

ABSTRACT

Know your customer (KYC) processes place a great burden on banks, because they are costly, inefficient, and inconvenient for customers. While blockchain technology is often mentioned as a potential solution, it is not clear how to use the technology's advantages without violating data protection regulations and customer privacy. We demonstrate how blockchain-based self-sovereign identity (SSI) can solve the challenges of KYC. We follow a rigorous design science research approach to create a framework that utilizes SSI in the KYC process, deriving nascent design principles that theorize on blockchain's role for SSI.

1. Introduction

Financial regulation has three primary goals: financial inclusion, financial stability, and market integrity [85]. To achieve the goal of market integrity, regulators have introduced several regulatory requirements into the financial sector, such as the Financial Action Task Force on Money Laundering (FATF) recommendations, which seek to prevent money laundering and the financing of international terrorism, as well as Basel III, in reaction to the global financial crisis in 2008 [3]. To remain compliant with this regulatory regime, financial institutions must perform in-depth due diligence to identify their customers and to understand the purpose of their activities, a process formally known as know your customer (KYC) [2], in which customers typically need to be physically present at the bank's branch or on a video call to provide personally identifying information, such as a passport or an ID card.

This process is problematic for banks, because it is cost-intensive, time-consuming, and inconvenient for customers [85]. Thus, there have been several attempts at improvement, mostly involving the digitization of particular process steps. For instance, some banks use their customers' analog proof of identity, such as passports, and create internally used digital customer identities to improve the process flow. However, this approach again suffers from inefficiencies, since it is error-prone, time-consuming [36], and highly repetitive [85]. The lack of shared standards and banks' reservations about sharing customer information with competitors also limit the reusability of a customer's KYC data at different banks [4].

A central utility that collects and provides identity-related data for an electronic KYC (eKYC) process, as in India or Australia, is often mentioned as a solution to the aforementioned problems [4,58,85], since it can reduce costs and significantly shorten KYC onboarding processes [59]. However, recent reports of leaks and misuses of personal data have lowered the confidence of both banks and customers in solutions that involve creating central data silos [74]. Moreover, there are jurisdictions in which such a centralized service run by the government is not feasible [64]. Generally, the fear that such a distinct service provider will aggregate significant market or political power impedes the establishment of a widely accepted centralized service provider [84].

Thus, both researchers and practitioners have identified blockchain technology as a potential solution to the latter problems. Blockchains can provide neutral platforms for digital cross-organizational workflows [30], mitigating the threat of market power aggregation. At the same time, blockchain technology enables digital trust through synchronized redundancy and therefore transparency, tamper-resistance, and enforcement of processes through smart contracts [65]. However, it is well known that blockchain technology's built-in transparency and append-only structure aggravates privacy-related problems [64]. Particularly, the European General Data Protection Regulation (GDPR) grants individuals the *right to be forgotten*, which means that they can demand that their private data be deleted at any time as soon as the purpose for their storage has expired. As data stored on a blockchain

[*] Corresponding author.

https://doi.org/10.1016/j.im.2021.103553
Received 30 July 2020; Received in revised form 20 October 2021; Accepted 24 October 2021
Available online 29 October 2021

practically cannot be erased, implementations such as Moyano and Ross's [51] where eKYC-related information is stored transparently on-chain, are not a viable solution.

As an alternative, one could think of depositing the KYC information in a standardized way at the one and only entity involved in each of its KYC processes – the customer. These considerations lead to the concept of self-sovereign identity (SSI), which seeks to establish holistic digital identity management on the paradigm that a user controls all their data and attestations, similar to today's analog identity management via a system of plastic cards in physical wallets. Yet, SSI is still strongly linked with blockchain technology because it requires a neutral platform that provides governance, standards, and essential public information to check the validity of attestations. This goal of an interoperable digital identity management system without a distinct central authority in control makes SSI very attractive for digitizing the KYC process.

A recent pilot in the UK that investigated the opportunities of SSI-based KYC found that an SSI-based "portable identity significantly improves both consumer experience and protection, while accelerating customer onboarding and reducing KYC and compliance-related costs for financial institutions" [42]. While research on the problem and approaches to SSI-based eKYC onboarding have recently emerged [71], they have not covered topics such as user orientation, coverage of the entire KYC process, or platform independence. Further, Soltani et al. [71] focused on implementing the principles of SSI without acknowledging that SSI is a tool to achieve an improved KYC process from the perspective of banks. Looking at SSI generally, in the related literature, blockchain's role in this context remains largely unclear. Thus, both research and practice need a generic and validated framework that guides the design of SSI solutions for entire eKYC processes and an overview of the resulting implications to assess the potential benefits and to learn how to leverage them. Further, we still lack generic (DPs) to guide the development of SSI solutions based on blockchain technology that can also be used in other sectors [44]. We seek to design a framework for an eKYC process built on blockchain-based SSI and to derive initial generic DPs. We develop and evaluate our framework in a rigorous design science research (DSR) approach, incorporating both existing theoretical knowledge and practitioners' perspectives through semi-structured expert interviews. Thus, we extend the literature on eKYC by providing a comprehensive architecture and process framework, discussing the roles of blockchain and SSI for eKYC, and producing generalizable knowledge on the design, opportunities, and challenges of blockchain-based SSI systems. The DPs we develop from our DSR suggest that blockchain's role in SSI should be more restrictive than is typically proposed in the literature in order to make systems scalable and compliant with regulations. We also guide practitioners on how to design the respective systems.

The remainder of this study is structured as follows: In Section 2, we present background knowledge on KYC processes, blockchain technology, and SSI that is necessary to understand the work that follows. In Section 3, we present our DSR method. In Section 4, we derive objectives for the eKYC framework and evaluate them through expert interviews. We present the framework, including the SSI-based eKYC architecture and process, in Section 5. Section 6 continues with the evaluation of the framework along the derived objectives. In Section 7, we discuss the findings, develop nascent DPs for blockchain-based SSI, and provide managerial and theoretical implications. In Section 8, we summarize our results, identify limitations, and provide an outline for further research.

2. Background

2.1. The KYC process and centralized attempts at eKYC

After the original FATF Forty Recommendations were drawn up in 1990, they were revised in 1996 to account for the latest money laundering techniques. These recommendations for anti-money laundering (AML) have been adopted by more than 130 countries and are therefore considered to be the international standards [66]. A key element of these recommendations is the KYC process. Financial institutions are urged not to open anonymous accounts or accounts with obviously fictitious names. In this context, due diligence is recommended to verify the identity of customers through independent, credible documents. The purpose of the business relationship must also be verified. Further, the KYC process should include ongoing monitoring of transactions to identify suspicious customer behavior [24].

KYC processes may differ, owing to countries' different regulatory requirements and the banks' specific requirements. However, some repeating core activities of the KYC process can be identified (see Figure 1). The process begins with the collection of data about potential customers to identify them. Government-issued documents such as ID cards, driver's licenses, or passports are preferred. Documents from other companies in the financial sector, as well as other documents relevant for the identification of persons, such as telephone or gas invoices, can also be used [52]. After a customer is identified and the identity data claims are verified, the bank checks whether the person represents a risk for the financial institution. This includes matching against a list of known terrorists, criminals, and politically exposed persons [2]. Further initial and ongoing measures follow to allow the bank to do permanent risk monitoring.

The process of initial verification and ongoing monitoring of activities must be repeated for each customer, and every customer must undergo this process again when opening an account with a new bank. Thus, the KYC process is very time-intensive and inconvenient for both customers and banks, resulting in poor customer experiences and fewer account openings. For instance, 89 % of surveyed customers did not have a good experience with the KYC process [76], and criticized the onboarding process because it was time-intensive and involved posting several documents. To avoid losing their customers and revenue opportunities, financial institutions must make essential improvements. Also, the overall market efficiency could benefit from enhanced competition owing to lower switching costs. Further, the high effort required for this process and the lack of automation of some manual steps result in high costs for the financial institutions. A survey of 800 financial institutions found that the annual cost for KYC per bank is approximately USD $60 million [76].

The primary focus and motivator of regulatory efforts toward KYC is the avoidance of money laundering through financial institutions. Failure to comply with regulatory requirements may further increase KYC process cost through considerable fines [51]. A major goal in KYC efforts for financial institutions is often, therefore, the avoidance of fines or loss of reputation, at preferably low ownership costs. However, some institutions also see KYC as an opportunity, since it enables them to better understand customers, identify their needs and behaviors, create customized products, and improve customer relationships, ultimately leading to higher company profits [66].

The key to simultaneously reducing compliance costs, preventing regulatory penalties, and harnessing new potential lies in the digitization and automation of processes and the resulting opportunities for data processing and analysis [45]. One often used approach to improve the KYC process is the digitization of analog ID documents, which typically involves some facial verification step by a combination of human and machine learning examination. One step further are approaches that seek to abandon analog documents altogether, which is why the term eKYC is often used here [14]. A sector-wide eKYC utility could avoid the repeated execution of the KYC process at different banks. These systems typically use biometrics such as fingerprints, iris scans, or facial recognition. The data are then stored on a smart ID card and online in a central database, together with personally identifiable information such as the customer's name, age, and place of residence. During the KYC process, the customer's biometric data are captured and matched against the data in the central online database.

An example of such a sector-wide eKYC utility is India's Aadhaar system. Indian citizens must provide various demographic and biometric

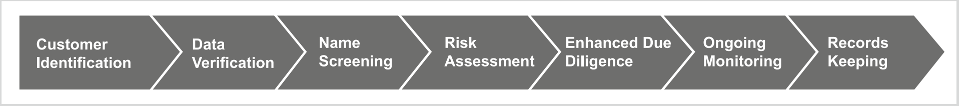

Fig. 1. The KYC process.

data [85], which are stored in a central database. The system has led to much faster onboarding times and to fewer losses from fraud and corruption [75]. However, several data breaches have raised questions regarding the privacy and security of the system [85]. Further, if operated by a public authority or heavily regulated, identity systems such as India's could be used by governments for mass surveillance without citizens' knowledge. On the other hand, on an international level or if operated by a private company, threats of monopolies and market power may keep banks from participating in such a system [84]. Thus, while it is a key step toward increasing process efficiency, the design of identity systems is critical for their success and acceptance, because this involves the management of highly sensitive data and misuse should be prevented. Centralized databases to store users' personal data are attractive targets for attackers who can steal large amounts of sensitive information. Consequently, they are challenging to secure [69]. Similar concerns regarding the creation of centralized service providers for non-competitive data arise in various further contexts beyond KYC. For the KYC procedure, digital identity systems must therefore be considered from the perspectives not only of efficiency and user experience but also privacy and security [58].

2.2. *Blockchain technology and decentralized approaches to eKYC*

Owing to the limitations and downsides of centralized platforms, banks have started looking for alternatives, one being distributed ledger technology (DLT). Key components of DLT are a peer-to-peer network where all data are replicated across multiple peers, and an associated consensus protocol operated by specific nodes to ensure the validity of state modifications (*transactions*) and to synchronize all replicas [28,40]. Authentication on DLT is conducted through public key cryptography, which allows one to participate in consensus or to interact with the network and authorize transactions. Distributed ledgers are resistant to crashes and even the malicious behavior of a small subset of nodes, making them a highly available and decentralized digital infrastructure. However, DLT also has considerable drawbacks concerning scalability and privacy, owing to the redundant operation of all transactions [40]. Blockchains[1] are a special case of DLT, and are probably the most widely used. The key characteristic of blockchain architectures is that transactions are batched into blocks, and each block of data contains the previous block's hash value. The blocks therefore form an append-only structure (*chain*) with the aim of establishing a tamper-resistant historical record [10].

Thus, blockchains can serve as a physically decentralized yet logically centralized source of truth for information, making them suitable for decentralized asset management [65]. Guaranteeing transparency and the enforcement of rules while ensuring the independence from a distinct node can be major advantages of blockchain solutions for cross-organizational workflow management [27]. Businesses and public authorities have realized DLT's potential for the digitization of their cross-organizational processes, leading to a large number of projects [12]. Considering the aforementioned generic properties of DLT, a blockchain-based neutral platform on which banks could collaborate on eKYC seemed very appealing, since this approach can eliminate the threat of monopolies. However, it aggravates privacy-related problems, since tamper resistance and redundancy imply not only that stored on-chain data are visible to all nodes but also that it is practically impossible to delete on-chain data [40,64]. It therefore does not make sense to store personal data on the ledger [21] and doing so contradicts regulation such as the GDPR, which includes the *right to be forgotten.*

This fact significantly complicates the conceptual integration of a DLT into the KYC process. Moyano and Ross [51], Biryukov et al. [9], and Norvill et al. [55], for instance, proposed writing a proof about the successful completion of the KYC process in the form of a hash value on a blockchain. In this concept, the de facto data are still stored in a centralized database operated by banks or a service provider. Once a bank customer has completed the KYC process, it will be sufficient for the customer to prove their identity using the hash value in the ledger. Although the efficiency of the process can thus be increased, central parties with full control of and access to the data are still necessary with these approaches, again causing the described security and privacy challenges. Further, challenges regarding the binding of cryptographic keys to customers as well as the management of permissions for exchanging customer data remain, while the benefit of using a blockchain is not yet clear, as a public key infrastructure and certificates based on digital signatures can provide tamper-proof evidence of a completed KYC process. [56] acknowledged the challenges of storing customer data on a blockchain in their development of a blockchain-based system for KYC to satisfy the requirements of initial coin offerings. Thereby, only the statuses of completed KYC processes are stored on a blockchain. However, in their design, the customers' identity data remain with a centralized provider specialized in KYC, and the protocol for exchanging data between the banks and the eKYC provider remains unspecified.

2.3. *SSI and its proposed application to eKYC*

Today, identification and authentication are usually carried out against a service provider using a username and password. The reason for the widespread use of this so-called *centralized* identity model lies in its simple implementation and in the full control of the service providers, who can minimize risks if no third party is involved for authentication. Users also benefit from the fact that they only have to pass on the information necessary for the context in question [15]. However, the increasing use of internet services has made this system inconvenient for users, since they have to remember the login data for each additional service, and manual input or repeated verification processes of attributes are necessary [22]. This leads to poor user experiences and security issues, as users tend to reuse passwords across many services. Moreover, service providers need to rely on the validity of the data provided by the customer, which can result in bad data quality and costs for fraud that cannot be traced back to a natural person. Service providers also usually store the data in large data silos – a popular target for hackers [59].

In an attempt to improve user experience, the so-called *federated* identity model was developed [46]. This concept allows for the use of digital identities for authentication and proof of attributes across organizational and system boundaries. An identity provider, such as Facebook or Google, manages users' digital identities and makes them available to relying parties. The fundamental prerequisite for this identity model is the establishment of a trust relationship between the identity provider and the relying party. Federated identity management improves user experience, since the users no longer have to remember a large number of user names and passwords, and only need a single sign-on [43]. However, from the perspective of privacy and security, such services are even more problematic than centralized systems [46].

[1] We use the terms blockchain and DLT interchangeably in this work

If privacy and security need to be improved, there must no longer be any central parties that have access to users' full digital identities and the associated data. Rather, control must be decentralized. By using public key cryptography, users can create their own identifiers – also known as decentralized identifiers (DIDs) – and prove control over them. Users can then append information to these identifiers. For contexts in which some attested attributes require confirmation, users can collect credentials from trusted authorities, such as government agencies, companies, or universities [73]. DIDs and the associated cryptographic keys, as well as credentials, are stored by users in so-called digital wallets, for instance on smartphones, computers, or in the cloud with a provider of their choice. Such a system is comparable to the physical credentials, e.g., plastic cards, we carry in our physical wallets [6]. Since users fully control their data, this approach has been called *self-sovereign* [1].

Such an approach requires open-source and open-standard technology [83]. Various implementations of SSI are possible and have been realized, but currently many commonly used implementations build on the DID standard being developed by the World Wide Web Consortium (W3C) [62]. A DID is always associated with a DID document that contains information such as public key material used to delegate and prove ownership and control of a DID [62], and to establish a secure (encrypted) communication channel with this DID. Besides the purpose of standardization, DIDs create a reference point for bilateral interactions that is portable across domains and does not require a centralized authority to register, resolve, update, or revoke the identifiers [71]. In this sense, DIDs are not strictly necessary for SSI, but provide functionalities that go beyond the mere capabilities of decentralized public key infrastructure (DPKI).

Credentials that provide cryptographic evidence of who created them and who they were created for are widely known as digital certificates. A new flavor, called verifiable credentials (VCs), is currently the subject of standardization efforts by the W3C [73]. Their validity and whether they have expired or been revoked can be verified without having to communicate with the issuer of the credentials, by checking the issuer's digital signature and a public yet privacy-preserving revocation registry. However, this approach requires an established trust relationship between a verifier and the credential issuer [53]. The decentralized approach regarding the reliable and trustworthy provision of public information that is necessary to verify VC data is enabled by the use of DLT. DLT acts as a *single point of truth* and thus as a generally acceptable and immutable location for the storage and management of information about standards, issuers of VCs (e.g., their public signing keys), and revocation status. DLT therefore provides a censorship-resistant storage facility for information that must be publicly available, without the need for a central entity such as a certificate authority [53]. SSI's key roles and building blocks are summarized in Figure 2.

Besides the security aspect, a widely acknowledged opportunity of SSI is enhanced privacy features [69]. One the one hand, by default, different identifiers, so-called pairwise DIDs (pseudonyms), can be used in different interactions. Global DIDs are required only for public entities that want to aggregate reputation or trust, such as credential issuers. Further, some implementations of VCs can prove the correctness of claims, such as the existence of the issuers' signature on the VC, without the need to reveal the value of the signature itself or all attributes that are attested on the credential. This significantly mitigates the correlatability of conventional digital certificates by means of their digital signature, and ultimately allows for enhanced privacy while still exchanging the information that is required to build the trust relationship that is necessary for interactions and business [18,31].

Since the concept places users in the center and leaves them in full control, some general challenges arise with SSI. First, appropriate measures must be taken to ensure user friendliness. Users must take care of storing the credentials and managing the keys themselves. So-called digital agents or wallets are used for this, either directly on an edge agent (e.g., a smartphone or laptop) or with cloud agents that can only be accessed by the user [77]. Cloud agents are helpful, since edge agents cannot guarantee permanent online availability [61]. Further, problems such as recovery in the case of device loss or theft must be addressed. Second, a governance framework is required to establish the possibility to gain trust in a large variety of issuers. Third, user authenticity must be guaranteed; i.e., sharing and selling credentials must be prevented [11]. However, various concepts that allow one to address these issues, such as initiatives to build governance frameworks [18] and the combination of biometrics, cryptography, economic incentives, and device coupling to create a strong bond between credentials and holders [31,32].

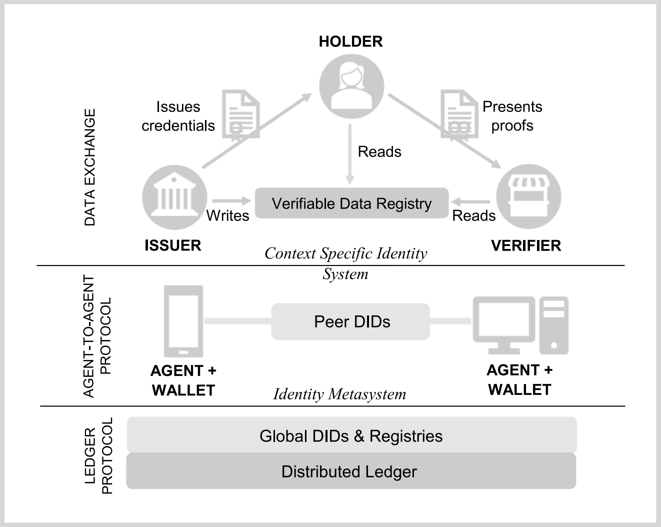

Fig. 2. Layers of SSI-based identity management based on [79].

In sum, SSI allows for highly decentralized management of personal identifiers and for credentials that are reusable across different contexts to be managed by the user from a single app [69]. In what follows, we investigate how this approach may help meet the challenges of the eKYC process.

3. Method

We followed a DSR approach. DSR was originally created to enable IS practitioners to find solutions to previously unsolved problems through a continual build-and-evaluate process. Its outcomes are IT artifacts, such as constructs, models, methods, or instantiations [34,47]. While some scholars argue that the IT artifact itself already contributes to research if it is novel and useful [7,29], two challenges in discerning DSR's research contribution remain: First, it is hard to determine what exactly a theoretical contribution in DSR is [29]. Second, it is hard to balance concrete, practical contributions to a rapidly changing technology environment and to provide a sufficient level of generalization for theory [7]. To address these challenges, we aim to contribute both an architectural design and a collection of processes as a concrete IT artifact [29]. To elevate this IT artifact for further theoretical discussion, we then derive DPs [29,34]. Thus, we aim to contribute nascent design theory in the form of operational principles [29].

For an IT artifact to offer a substantial contribution to IS research, it must address a relevant business need [34], which can result from the persons, organizations, or technologies used in an environment. As argued in Section 2.1, the enhancement of the KYC process represents such a business need. However, an IT artifact must also be applicable in the corresponding environment [34]. To ensure rigor in the design process, the construction of the IT artifact needs to build on existing foundations from previous IS research [82]. Also, existing methodologies should be used to evaluate the created artifact [34]. The KYC framework here is based on related work that aims to improve the KYC process using digital technologies, the technical and theoretical foundations of KYC, DLT, and SSI, and the requirements and expertise of practitioners in said areas.

We employ the frequently used and widely accepted [63,68] DSR process model of Peffers et al. [57] to facilitate the development of a relevant IT artifact created by a rigorous method. Our process has six steps arranged in sequential order (see Figure 3) and incorporates an iterative research procedure by design [57]. The process typically starts with the identification of a research problem with practical relevance. Indeed, as illustrated in Section 2.1, our examination of the current KYC process reveals challenges such as low process efficiency, security challenges, poor user experience, and data protection concerns.

Next, we defined solution objectives to address the stated challenges and to create a meaningful artifact. In line with DSR, the insights gained from the build-and-evaluate process must be generalizable and therefore applicable in more generic settings [37]. Also, the design artifacts should result in profound disruptions to traditional ways of doing business [33]. Recent research into DSR encourages researchers to build their work on prior DSR within the respective domain [82]. We derived solution objectives by studying the related literature and regulatory requirements, both for the KYC process and for digital identification and authentication, resulting in six main objectives for the KYC framework and several requirements for each main objective. Based on these objectives and on theory, we design and develop an SSI-based eKYC framework in the next research process step. Phase 5 comprises evaluation, which is necessary to test whether an artifact achieves the purpose of its creation and to prove this achievement using rigorous methods [80]. The evaluation phase also helps one to better understand the problem at hand and thus to realize improved outcomes [34].

There is no unique path regarding evaluation, since the best approach depends on both the underlying problem and the artifact [57]. Our evaluation had several iterative evaluation steps, starting ex ante with the formative evaluation of the design objectives through interviews with experts [72,81]. We conducted six additional ex post interviews to summatively evaluate our framework by demonstrating it to the interviewees and incorporating their feedback. The evaluation of the framework was designed to assess its functionality, accuracy, reliability, fit with the organization, and utility [34]. We then applied a criteria-based evaluation concerning whether the derived solution objectives were met, since evaluation criteria for an IT artifact must themselves be determined for the particular environment [47]. To elevate the implicit knowledge contribution in our IT artifact to more abstract and generalizable knowledge allowing for theoretical discussion [29], we then developed nascent DPs for blockchain-based SSI, as this technical approach is both novel and increasingly discussed, though no general DPs currently exist in the literature. Finally, we shared the findings of our research with the relevant audience [34]. The applied DSR process was iterative and partly in parallel, since the evaluation phase's results have reshaped the created artifact [8].

Qualitative interviews, as used for our evaluation cycles, are a frequently used method in IS research, since they are suitable for generating rich data [54]. We conducted semi-structured interviews so that we could react flexibly to the interviewees' answers and ask appropriate follow-up questions [38]). We involved experts on KYC and SSI to reflect opinions from the perspective of practical applicability in existing settings and bank structures as well as opinions regarding technical maturity and feasibility. Also, we took care to avoid an elite bias by representing the voices of executives of different corporate levels [54]. Further criteria for the selection of the experts included ample knowledge of their disciplines and intensive experience in their daily work, as well as the ability to provide detailed information on their field of expertise [50]. A detailed overview over the interviewees appears in Table 1.

We recorded 320 interview minutes (an average of 35.6 minutes per interview). The interviews were recorded, transcribed, and later analyzed using MAXQDA software. For data analysis, we used both open and axial coding [67]. Starting from the initial concepts originally derived in the open coding round, categories were formed. Categories are "higher-level concepts under which analysts group lower-level concepts that then become its subcategories" [16, p. 220]. During this first coding round, we created 30 categories and 300 subcategories; in the second, we used axial coding to build subcategories. Thus, the data that were split up during open coding were reassembled to summarize the categories on a more abstract level [13,16,67].

Table 1
Overview over the Interviewed Experts.

Episode	Expertise	Id	Role	Background	Type
1	KYC	A	Project Manager KYC	Strategic Analysis and Research, Banking	Phone Call
1	KYC	B	Sales Director	Building Society, Banking	Video Call
1	SSI	C	Identity Engineer	Innovation Consultant	Video Call
2	SSI & KYC	D	Executive Director	SSI Start-up Founder, Banking	Video Call
2	SSI & KYC	E	Project Manager	Banking	Phone Call
2	SSI	F	Senior Developer	Computer Science	Video Call
2	SSI	G	CEO	SSI Start-up Founder	Video Call
2	KYC	H	Sales Executive	Banking	Video Call
2	KYC	I	Sales Director	Banking	Video Call

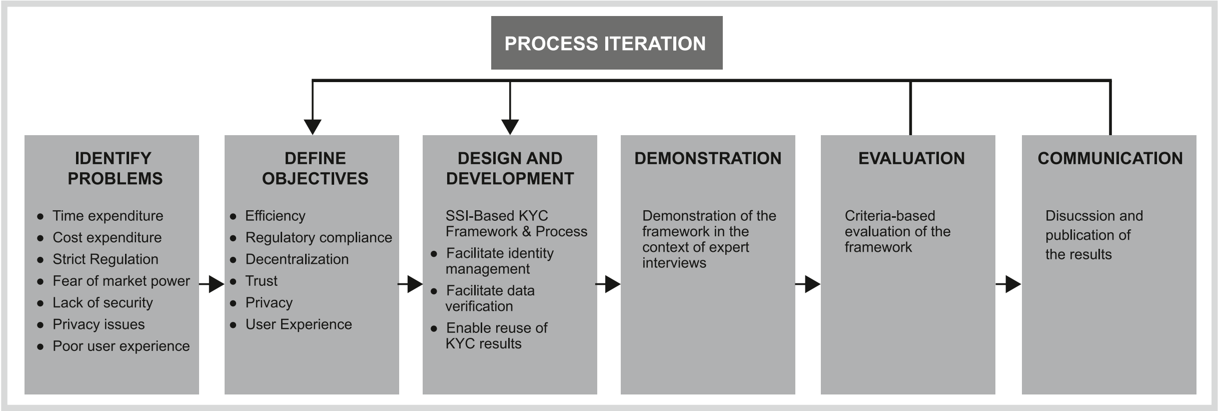

Fig. 3. Our applied DSR process, following [57].

4. Design Objectives for the eKYC Framework

4.1. Structuring of design objectives

To comprehensively address the challenges of the KYC process (as identified in Section 2), stage 2 in our DSR process involved the derivation of objectives to be met by a useful SSI-based eKYC framework. We derived these objectives from the literature on the KYC process, KYC-related regulatory requirements, and three formative interviews with experts. Thus, we aimed to align with DSR by incorporating prior research [82] and incorporating real-world business needs [34]. We identified six main objectives and associated requirements. In what follows, we explain and justify them.

Objective 1: Efficiency

The high cost and human resources involved in carrying out the current KYC process strongly challenges financial institutions. Financial institutions offering fast and convenient verification of identity documents are more attractive from the customer's perspective, can reduce costs, and can help a company gain a competitive advantage [36]. To allow for increased process efficiency, we derived three requirements that had to be satisfied to overcome the existing challenges. The *end--to-end digital processing of relevant documents (R 1.1)* is a prerequisite for automating process steps and reducing friction [4]. Further, in the current KYC process, many steps involving the validation of data, such as checking whether an ID document's validity has expired, are conducted manually [85]. Thus, the de facto *automation of manual processes (R 1.2)* is another key requirement. Further, Moyano and Ross's [51] interviews with five senior banking executives revealed a need for interbank collaboration; this was confirmed by Experts A and B, who stated that banks would be ready to collaborate on resource-intensive KYC. Currently, however, the main barrier to such cross-organizational processes is the lack of a suitable non-proprietary IT infrastructure. Thus, a *standardized exchange of eKYC documents (R 1.3)* is crucial to allow for the efficient integration of eKYC checks that have been conducted at other institutions.

Objective 2: Regulatory compliance

Compliance with regulations is a key objective of the KYC process [56]; derived from the overall goal of avoiding money laundering, it is one of the main reasons why the KYC process exists at all. Our literature study revealed that the *Money Laundering Act (MLA) (R 2.1), GDPR (R 2.2)*, and *electronic Identification, Authentication, and Trust Services (eIDAS) (R 2.3)* are particularly relevant regulations for a digital KYC process [4]. While these apply within the European Union (EU), there are similar regulatory requirements in other jurisdictions worldwide. The European requirements are considered particularly strict, which is why we decided to apply them here.

The MLA provides banks with specific requirements regarding the identification of customers and the storage of their records. The banks are also required to determine and document the risk in relation to their customers. The GDPR applies to the processing of any data regarding natural persons, but not to legal entities, and poses requirements such as privacy by design, portability, the right to erasure, transparency, purpose limitation, data minimization, accuracy, storage limitation, information integrity, and confidentiality. Further, digital KYC processes involve the customer's identification and a check of the authenticity of the involved documents, and the 5th EU AML Directive accepts electronic ID systems that comply with eIDAS as a legitimate means of identification for KYC procedures. eIDAS imposes requirements on these electronic means of identification, such as compliance with certain security levels (level of assurance) and the cross-border interoperability of systems.

Objective 3: Decentralization

As argued in Section 2.1, silos of customer data are an attractive target for attackers. Securing valuable information is costly and not the core business of banks, and mistakes can have severe consequences concerning reputation, fines, or both. Recent data breaches that revealed sensitive customer data stored in central data silos have significantly reduced confidence in their respective architectures [59]. To avoid comparable data breaches, a viable solution for an improved eKYC process must therefore *avoid central storage of customer data (R 3.1)*. Further, banks do not want to risk becoming dependent on a centralized eKYC service provider. Thus, the system must be constructed to *prevent lock-in effects (R 3.2)* that could result in the aggregation of market power. Decentralization of both data storage and workflows is therefore one key objective of the new eKYC architecture.

Objective 4: Trust

A key goal of banks is to make eKYC documents reusable in registrations of a customer at different banks. If banks do not comply with the regulations, there can be heavy fines, so it is important to establish trust in the KYC process and the integrity of its documentation at other banks. Thus, *acceptance of KYC documents attested by other banks (R 4.1)* is required. The documents must be tamper-proof, so a further requirement is that *validity checks (R 4.2)* of these documents are feasible. Another often disregarded requirement for a complete trust chain is that sharing or selling KYC documents among customers must be prevented. This can be particularly difficult if the eKYC process happens remotely and lacks interaction with an employee of the bank. The customer needs to be able to convince the bank that the KYC-related documents that they present were not stolen, sold, or shared. We call this requirement *authenticity checks (R 4.3)*, meaning that the identity of the customer and their connection with the documents must have a high level of assurance even if the customer is not present at a branch and no video call is held.

Objective 5: Privacy

Protecting customers' privacy is a key feature of an eKYC process. Facing an increasing number of data leaks, customers are aware of privacy issues, and delivering a privacy preserving solution may increase the solution's acceptance. An essential and fairly universal principle in this context is *compliance with the "need to know" principle (R 5.1)*: Only the customers themselves and entities relevant to the KYC process must have access to customers' personal data. This is also a general recommendation for information systems from a security perspective [35,48]. Further, not only the parties involved in the KYC process but also the de facto data that are exchanged should be restricted to what is necessary, because digital data are much more comprehensive and easier to collect and abuse than their analog counterparts [4]. We call this requirement *data minimization (R 5.2)*.

Objective 6: User experience

From the users' perspective, although privacy is a nice feature that can be used for marketing purposes, the most important objective is seamless user experience [39]. The eKYC process must be convenient, so that customers are not discouraged from registering at the new bank. It is only when the eKYC process is fast and simple for the customer that it can provide high security and acceptance [20]. Thus, we made *low complexity (R 6.1)* a major requirement for user experience. Further, the variety of devices on which a customer can perform the eKYC process must be respected. Mobile phones are often the customers' preferred option, but support for web apps is also necessary in many circumstances. Thus, the availability of *different user interfaces (R 6.2)* is important. The user experience should also include exception handling, for instance, if a device that stores the customer data is lost or stolen. In this case, either there must be a built-in recovery mechanism, or the customer must be able to ask for rapid support. This is very difficult if no central third party is responsible for the whole process. Thus, we also added such *backup, recovery, and support (R 6.3)* features to our requirements.

4.2. Evaluation of the design objectives

We discussed the current problems of the KYC procedure and our derived objectives with two KYC experts and an SSI expert. The interviews sought to evaluate the identified design objectives concerning relevance and completeness. The KYC experts worked in different companies and held different positions, so that the objectives could be viewed from different perspectives. Additional information on the interviewees appears in Table 1.

Expert A confirmed the relevance of the derived objectives and their associated requirements. Owing to the increasing expenditure on personnel and technology, the process's efficiency is indeed a crucial goal for banks. He stressed the importance of end-to-end digital processing and advocated interbank cooperation in the KYC process, but identified trust problems here, both between the banks and concerning customer trust in the confidentiality of their data. According to him, the protection of customer privacy is also crucial. Further, he affirmed the relevance of increasingly strict regulations and the need to comply with them. For instance, customer data must be stored by banks for at least five years. The expert also confirmed the necessity of including further MLA requirements.

Expert B also described process efficiency as the most crucial factor, to ensure cost and time savings. The challenges apparently lie particularly in the high number of manual process steps. This expert emphasized the importance of automation and digital processing of documents. He also confirmed the importance of protecting privacy. Sensitive handling of customer data is necessary, and this must not be passed on to third parties, not even to cooperation partners. Like Expert A, he noted the increasing importance of regulation and the need to comply with it.

Expert C emphasized the importance of a good user experience, since many users will not focus on the systems' functional details. During the implementation phase, special care should be taken to ensure that the system is as intuitive as possible. Asked about the architectural perspective, he mentioned backup and recovery capabilities through cloud storage as a building block for user friendliness in case of data theft or loss. Expert C also confirmed the importance of the GDPR and eIDAS. According to him, there is still room for interpretation in the GDPR, for instance regarding the role of encrypted or hashed personally identifying data. He advised proceeding from the strictest possible interpretation of the GDPR. He stressed that, on a distributed ledger, data cannot be deleted. A key challenge to the acceptance of KYC documents attested by other banks, he spoke of the necessary establishment of a trust relationship between the banks. However, he argued that connecting the eKYC architecture to the eIDAS infrastructure could be a solution to this problem.

In sum, at least one expert emphasized each of the design objectives, and the experts generally considered the associated requirements useful to evaluate an eKYC framework from a bank's perspective.

5. A Framework for eKYC Processes Built on Blockchain-Based SSI

5.1. The SSI-based eKYC architecture

Based on the related work presented in Section 2, we designed a decentralized architecture that seeks to address the challenges of the KYC process. The study by [51] motivated a decentralized design of eKYC to allow for inter-bank collaboration. However, the proposed system seems critical from a data protection perspective, considering the privacy-related challenges of storing customer data, also in encrypted or hashed form, on a blockchain system. We also noticed that the mechanism they presented does not enforce the alignment of incentives, since the integrity check only requires a local read operation on one node. Thus, while we appreciate the background they gave on the necessity of a reusable KYC and a non-centralized solution, we found the design of the SSI-based framework proposed by [71] more appropriate. Nonetheless, we generalized this solution by considering both the initial onboarding to receive the first KYC document and how an existing SSI ecosystem and regulation such as eIDAS integrates with and further strengthens eKYC. We also added an investigation of the framework's practical feasibility by rigorously evaluating our SSI-based framework concerning the technical, economic, and legal requirements. From a technical perspective, we abstracted from their solution based on Hyperledger Indy to a more generic perspective on SSI and extended their findings by rigorously evaluating the design.

Following [71] and the general approach of blockchain-based SSI, the proposed eKYC architecture involves three primary parties: the customer (holder), a bank (verifier), and an issuer (the same bank, another bank, or any third party trusted by the verifying bank, such as a government agency). Credentials from different issuers can also be conjugated, because in a verifiable presentation (VP), attributes attested in different VCs can be combined [73]. For simplicity, we assumed one issuer. The customer is the KYC subject and defines the center of the architecture (see Figure 4). Customers manage their digital identity through user agents by creating and storing DIDs and cryptographic keys in their digital wallets, collecting credentials, creating backups, and managing permissions. It is possible to interact with agents on various devices, such as smartphones or laptops. At all times, the customers have full control over their data and particularly over KYC-related documents, represented by VCs. While traditional certificate-based approaches (e.g., X.509 certificates) need to be shown fully to the verifier in order to check the signature's validity, the VC standard [73] and related implementations such as Hyperledger Aries allow for creating proofs from the VCs, convincing the verifier that certain claims extracted from the VC are correct without the need to exchange the full VC. This builds on cryptographic constructions such as anonymous credentials introduced by [11]. In our case, the VPs contain proofs of the validity of the attributes that need to be revealed during the KYC process.

To facilitate the redundant storage of credentials and easier user access to the SSI documents, as well as to enable secure communication with other entities, the framework employs cloud agents and wallets. The permissions for carrying out identification activities differ between edge and cloud instances. While edge agents and wallets are usually granted full access to an individual's data, the user should use cloud agents/wallets primarily for redundant storage and communication with other entities. A blockchain serves as a neutral infrastructure for storing publicly verifiable information. It is used to hold VC issuers' public signing keys and other institutional information. Further, schemas of KYC VCs are stored on-chain to allow for public verification. Also, publicly available revocation registries are stored on a blockchain to allow for public checks of privacy-preserving revocation information. Which credentials are accepted in the KYC process may be defined by each bank, depending on its requirements and trust relationships. The combination of eIDAS and DIDs could allow for qualified digital signatures that comply with eIDAS [23]. Credential issuers use institutional agents that are explicitly designed for creating credentials. Besides issuing credentials, these agents perform identification activities such as checking credentials for integrity and direct communication with the customer that is relevant during and after the KYC process. It also has an interface to name screening services, the bank's risk engine, and customer monitoring. The financial institutions are obliged to store data about customers, for which they use separate storage.

5.2. The SSI-based eKYC process

In accordance with the generic procedure of KYC processes, we split the proposed SSI-based eKYC process into three parts: (1) customer identification, data verification, and identity authentication; (2) name screening, risk assessment, and enhanced due diligence; and (3) ongoing monitoring and records keeping. The first part involves three scenarios, depending on the customer's status in the KYC process.

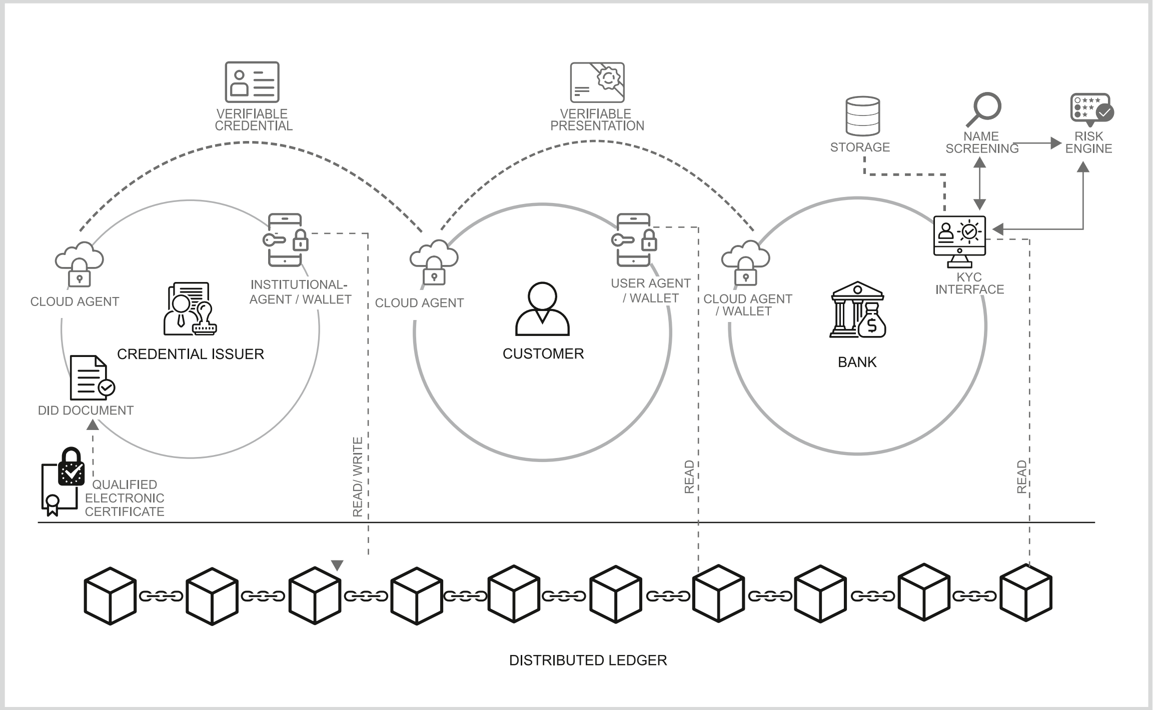

Fig. 4. SSI-based KYC architecture (based on [23,51,61,62,71,73]).

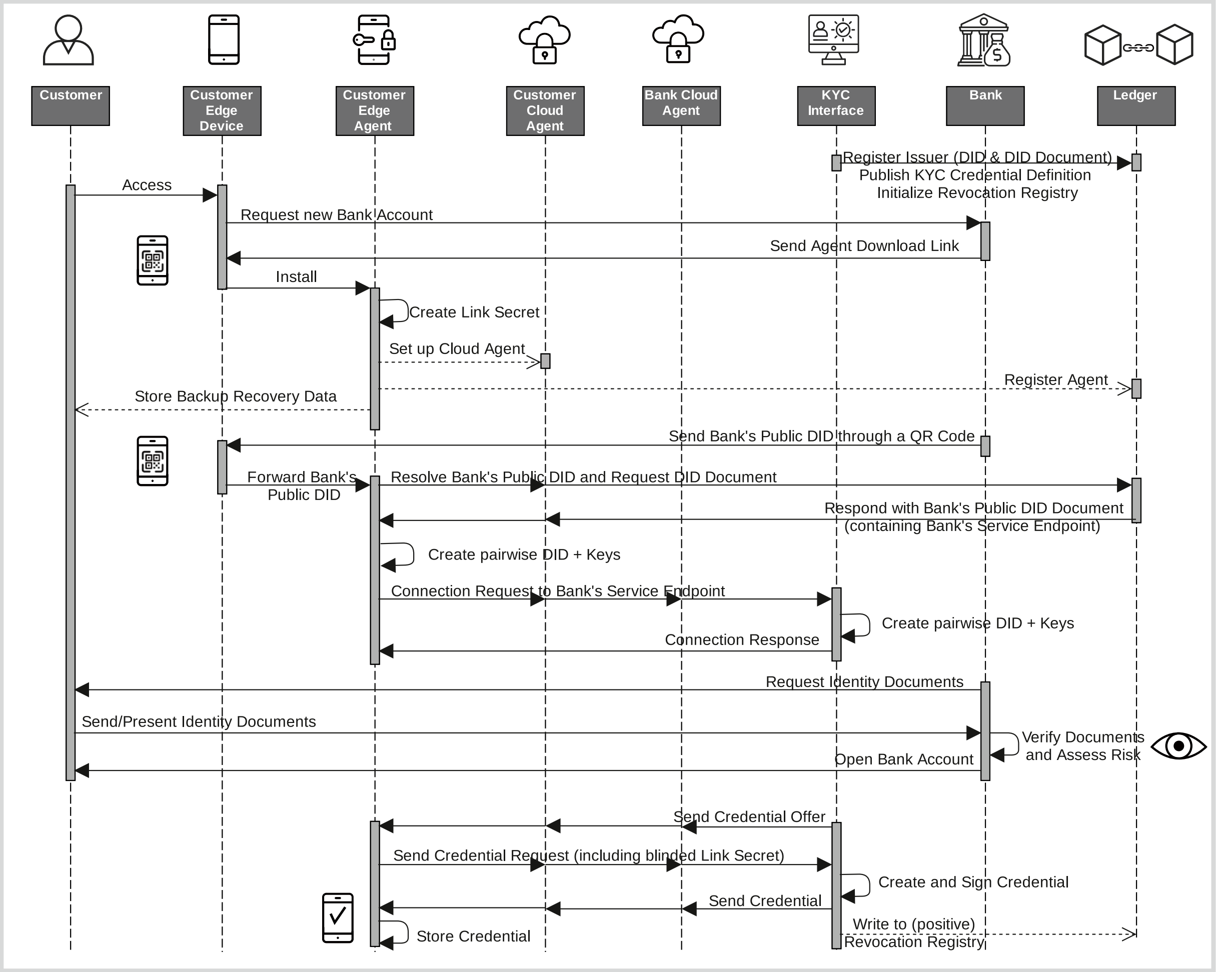
Customer
Customer Edge Device
Customer Edge Agent
Customer Cloud Agent
Bank Cloud Agent
KYC Interface
Bank
Ledger
Register Issuer (DID & DID Document)
Publish KYC Credential Definition
Initialize Revocation Registry
Access
Request new Bank Account
Send Agent Download Link
Install
Create Link Secret
Set up Cloud Agent
Register Agent
Store Backup Recovery Data
Send Bank's Public DID through a QR Code
Forward Bank's Public DID
Resolve Bank's Public DID and Request DID Document
Respond with Bank's Public DID Document (containing Bank's Service Endpoint)
Create pairwise DID + Keys
Connection Request to Bank's Service Endpoint
Create pairwise DID + Keys
Connection Response
Request Identity Documents
Send/Present Identity Documents
Verify Documents and Assess Risk
Open Bank Account
Send Credential Offer
Send Credential Request (including blinded Link Secret)
Create and Sign Credential
Send Credential
Store Credential
Write to (positive) Revocation Registry

Fig. 5. UML Diagram: *Completely New Onboarding*

Customer identification, data verification, and identity authentication

The KYC process starts with customer onboarding, where three cases can be distinguished. The first case is *completely new onboarding*, where the customer has neither an SSI agent/wallet nor VCs that confirm a completed KYC process. The second case, *fast onboarding*, is possible if the customer already has an SSI agent/wallet with corresponding VCs that attest to the prior completion of a KYC process. Third, we discuss a simplified case we call *new to KYC*, where customers already have an SSI agent or wallet and some VC from other contexts that contain identity-related information trusted by the verifying bank (or that the bank is allowed to trust from a legal perspective), but do not yet have VCs that demonstrate the completion of the KYC process at some institution.

We present the first case, *completely new onboarding*, in a UML sequence diagram (see Figure 5). To enable SSI-based onboarding, as illustrated in [71], banks must conduct a one-time bootstrapping process in which they first store a public DID and an associated DID document in a distributed ledger. This DID document may contain service endpoints of the bank, e.g., for obtaining customer services or conducting the eKYC. The bank will also publish a so-called credential definition, which may be derived from an agreed-on schema/template that contains the attributes that should be attested in a credential, and a revocation registry. All this information is meant to be publicly readable and contains cryptographic information that allows banks (verifiers) to check the validity of VPs that use an associated VC, and customers (holders) to conduct proofs of non-revocation. Thus, these can be stored on the blockchain layer, and no GDPR-related problems are to be expected.

After this initial setup, the bank is ready to perform customer onboarding processes. While [71] presented a (slightly less detailed) sequence diagram for customer onboarding, it involves reading from and writing to the blockchain more often than technically necessary. According to our interview with a co-author of the W3C DID standard, it suffices and is preferable from a privacy perspective to have a peer DID for the customer. We also discuss in detail the implications of the design for the bank and the customer, for instance, related to binding, revocation, and backups. *Completely new onboarding* starts with a bank customer who either visits the bank's website with their smartphone or laptop or physically arrives at a local bank branch to open a new bank account. Since the customer has neither an SSI user wallet nor the necessary KYC credentials, the bank recommends or offers a user wallet and provides the customer with a corresponding download link. Customers can download any digital wallet of their choice that supports the public DID, peer DID, and VC standards. It stores credentials and keys and is secured by a password or biometrics. The bank could further offer an edge agent in encrypted form in the cloud as capability for backup and recovery. The user wallet creates a new DID and some associated keys required for encryption and stores them in the wallet. At the start, the user also creates a so-called link secret, which will later be used to tie different credentials together in a VC and thus provides a means to prevent selective credential sharing. However, as long as all credentials contain the customer's name or another strongly binding attribute that needs to be revealed in the VP, it is easy to prove that they belong together also without a link secret.

The customer can now use their newly generated DID to establish an end-to-end-encrypted (secure) connection to the endpoint that the bank offers for the eKYC. The bank could also provide this information by submitting a QR code to the customer (e.g., via e-mail). The customer scans this QR code with their wallet app and thus connects to the bank's public service endpoint. The bank's service behind this endpoint now also creates a new pairwise DID as well as a key pair that the bank will use exclusively in this relationship, and sends a connection request to the customer's service endpoint, its cloud agent, which forwards the connection request to the customer's wallet app. This connection request contains the bank's pairwise DID, the public key used by the bank, and the service endpoint at which the customer can contact the bank, and could also involve a proof that the pairwise DID has in fact been authorized by the bank (e.g., through a VP in which the bank reveals its legal identifier that has been certified by a reputable public institution). In turn, the customer's digital wallet checks the connection's authenticity and creates a pairwise DID and keys for the relationship with the bank. Next, it sends a connection response to the bank's cloud agent/wallet, which forwards it to the bank's KYC interface. Now an end-to-end-encrypted connection exists between the bank and the customer, which can be used to securely exchange messages, public keys, VCs, and VPs. Since the customer does not yet have VCs, the customer's identity must first be verified. The customer sends the necessary analog identity data to the bank, either by traditional means or – if feasible – in scanned form via e-mail or the just-established connection. If the customer opens an account in a bank branch, the documents can also be verified directly there.

After the data have been verified and the customer's identity has been confirmed, the bank can send a credential offer to the customer's edge user agent via the established connection. This credential offer contains a preview of the data that will be attested, the credential issuer information, an expiration date for the VC, and information regarding credential revocation. The customer then accepts the credential offer and sends it to the bank, containing the link secret in blinded form.[2] However, the customer only has to create the link secret once and can later reuse it for their other VCs. The bank includes the blinded link secret in the attributes attested in the VC and sends the VC to the customer. The credential could support selective disclosure. That is, in any VP, the customer can include only the attributes attested by the VC that are necessary for the verifier, and combine claims from different VCs into a VP.

If the issuer wants to support revocation and has bootstrapped a revocation registry, the VC also contains information on how to check its revocation status. The credential issuer can then revoke credentials by updating a revocation registry in the distributed ledger. The bank, for instance, can use this mechanism to invalidate a credential that turns out to be wrongly issued. Notably, it is only through the additional information regarding revocation in the credential that the customer can make sense of the information in the public revocation registry and create a proof of non-revocation within a VP that contains attributes from this VC. Since the credential is never revealed, but only proofs are derived from this, this likely makes public revocation registries compliant with the GDPR.

Going beyond [71], we present in detail how the reuse of a KYC process works. This *fast onboarding* process also begins with a bank customer visiting the bank's website or a local bank branch to open a new bank account. The customer states that they already have an SSI user agent and VCs. In the case of opening an account online, the bank sends the customer its bank public DID, for instance by means of a QR code that can be scanned by the customer's wallet app. The channel by which the customer receives this information must be trusted, as the customer does not know the bank's DID in advance. Using an identity infrastructure such as eIDAS, the customer could check whether they are really communicating with the corresponding bank by verifying an eIDAS certification on the bank's public key in the DID document. The customer's user agent can then identify the distributed ledger that stores the DID document associated with the DID and can query this ledger for the DID document. The user agent uses the DID document to identify the bank's eKYC-related service endpoint. The customer's user agent now creates a pairwise DID for this relationship and the corresponding keys and sends a connection request to the bank. The connection request also contains the customer's pairwise DID and the public key used. The bank then creates a pairwise DID and corresponding keys, and sends the

[2] To be precise, the blinded link secret is a cryptographic commitment, i.e., the hash of the link secret and some one-time random number. Thus, while the blinded form will differ in each credential issuing process, the customer (holder) can still prove that different commitments originate from the same link secret, without revealing the link secret itself, in a zero-knowledge proof (ZKP).

customer a connection response, including the pairwise DID and public key.

After establishing the secure connection, the bank sends a proof request for conducting *fast onboarding* KYC. This request contains a random nonce to prevent replay attacks and specifies which data the customer must transmit to the bank, and restrictions on when to accept the VP. This includes a specification of issuers (credential definitions) and schemas that are accepted for the VCs used, and whether there is a need for a proof of non-revocation, including a timestamp of the revocation registry that the customer should refer to in creating this proof if a proof of non-revocation is demanded. The customer's edge agent automatically searches for VCs stored in the customer's digital wallet that match these requirements, updates their local revocation registry through a query if it has not been cached before, and creates a VP that it sends to the bank. The bank can now cryptographically verify the claims, which may also involve reading from revocation registries and other information regarding credential definitions unless sufficiently timely local data from previous queries are cached.

The bank can now cryptographically verify the proof, which involves checks that the digital signatures of issuers were on all attributes involved in the VP, that none of the attested attributes came from a revoked VC, and that all the VCs involved were issued to a commitment of the same, common link secret. After the proof has been verified, the bank account is opened. This whole process can be highly automated and is completed in a few seconds. The secure channel established between the bank and the customer based on pairwise DIDs can be used in the future to exchange further documents and to communicate securely and reliably. Based on the exchanged keys, a unique authentication of the involved entities is thereby possible. This is important when dealing with digital identities, since the bank must ensure that it communicates with the same person over time [36].

The third case is *new to KYC*, where the customer already has an SSI user agent and maybe even identity-related VCs, but does not yet have a VC that is accepted by the banks during the KYC process. Thus, *new to KYC* is a combination of *completely new onboarding* (Figure 5) and *fast onboarding* (Figure 6). While the construction of the pairwise DID relationship between the bank and the customer corresponds to the fast onboarding process, the transmission of the analog ID documents and the possibility of getting a KYC credential from the bank corresponds to the process of *completely new onboarding*. However, the customer could first check, through a proof request, whether only a subset of ID documents is necessary because some digital identity proofs are already in their wallet. In addition to the option in which the customer opens an account online, it is also possible to open an account directly in a bank branch by using a QR code to receive the bank's service endpoint and have the analog ID documents checked directly in the bank.

Name screening, risk assessment, and enhanced due diligence

After the identity data has been exchanged and cryptographically verified, the name screening service runs in the background of the bank's IT system to check the data against publicly known blacklists regarding terrorism, illegal money laundering activities, politically exposed persons, and negative press. The result of the name screening service is then fed directly into the risk engine, which uses this and other information to classify the customer into a risk class. The risk engine then calculates a risk score and classifies the customer into low,

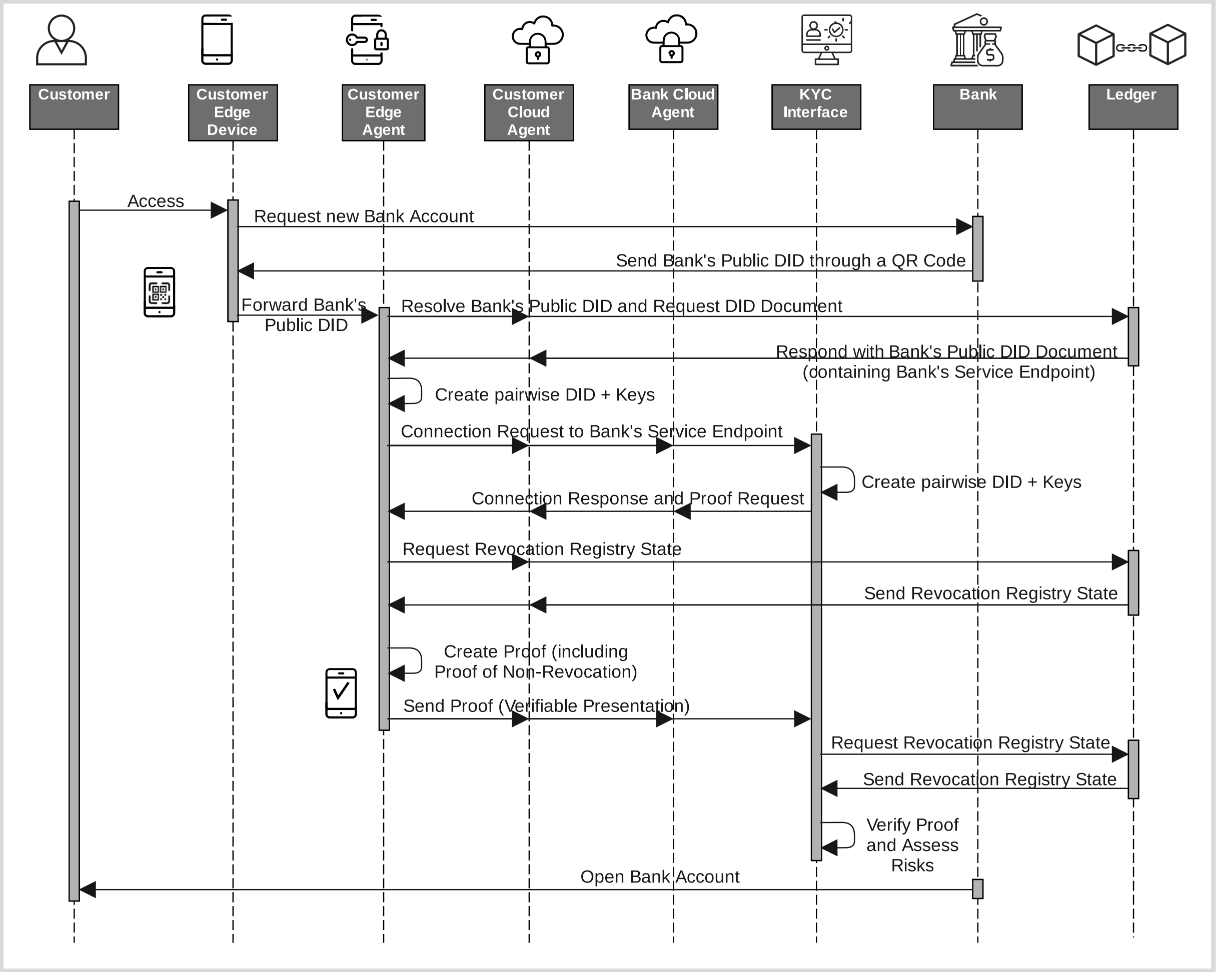
Customer
Customer Edge Device
Customer Edge Agent
Customer Cloud Agent
Bank Cloud Agent
KYC Interface
Bank
Ledger
Access
Request new Bank Account
Send Bank's Public DID through a QR Code
Forward Bank's Public DID
Resolve Bank's Public DID and Request DID Document
Respond with Bank's Public DID Document (containing Bank's Service Endpoint)
Create pairwise DID + Keys
Connection Request to Bank's Service Endpoint
Create pairwise DID + Keys
Connection Response and Proof Request
Request Revocation Registry State
Send Revocation Registry State
Create Proof (including Proof of Non-Revocation)
Send Proof (Verifiable Presentation)
Request Revocation Registry State
Send Revocation Registry State
Verify Proof and Assess Risks
Open Bank Account

Fig. 6. UML Diagram: *Fast Onboarding*

standard, or high risk. Depending on this result, further checks may be necessary before the bank opens the account. Since, in contrast to analog ID documents, VCs are much harder to forge, it suffices to request a minimum amount of information at the start of the relationship. Depending on the risk assessment's result, additional checks may become necessary later. To mitigate risk, the bank can use the previously established secure communication channel to request additional documents and information, such as an income statement or the reason for opening the account. As illustrated, such additional documentation could again be provided in analog form or by using VCs that are already in the customer's wallet – for instance, an income statement issued by an employer that the verifying bank trusts – and deliver an associated VP. Once the customer's verification is successfully completed, the bank can open the account.

Ongoing monitoring

Once the account has been opened, the risk engine checks the customer's ongoing transactions during the business relationship, compares these to the expected transaction volume, and checks the transactions for suspicious transaction patterns. Further, the risk engine regularly checks whether the expiration dates contained in the VPs have expired; these may even occasionally trigger a new proof request to the customer through the secure connection to ensure that none of the customer's VCs that were used for KYC have been revoked. The customer then only has to press a confirm button to deliver a new VP. Thus, a manual check of the identity documents is no longer necessary. If it turns out in a periodic refreshment that a customer's VC has in the meantime been revoked (which may just be because of a change of address or a successive reissuance of an ID card) or that the transaction behavior is abnormal, the risk engine reassesses the risk and proposes measures to mitigate these risks if necessary. The bank can then also request an updated version of the customer's VCs or further documents. This could even be extended to offering the customer an option to automatically send updated versions of their VCs (e.g., if the address on the customer's government-issued identity VC changes) to the bank after the KYC process, so that no more manual activities by the bank and the customer are necessary to keep the data up to date.

Record keeping

The SSI concept theoretically allows a bank not to store personal data about its customers at all. The data are solely stored in the customer's digital wallet, and it is very easy to request data when needed and convenient for the customer to provide this information. However, depending on the specific regulations, the banks may be obliged to store their customers' data for a longer period in order to be able to unambiguously determine the person's identity in the event of suspicious or illegal conduct. Therefore, the bank also stores the data in a local database in the redefined KYC process. However, the bank may still manipulate some data. A benefit of ZKP-oriented VCs is that the VP could be made either repudiable or non-repudiable, ensuring tamper-proof documentation where required and supporting customer privacy even in the case of hacks if no auditability is required or sensitive information such as income is involved [31].

6. Evaluation

We now report on a summative, criteria-based evaluation of our proposed framework with interviews with experts (as described in Section 3), evaluating each of the objectives derived in Section 4 and their associated requirements in detail [47].

Efficiency

According to Experts E and I, the SSI-based KYC process presented in the framework has the potential to solve the inefficiencies in the existing KYC process. This can mainly be achieved because the framework involves fully digital cryptographic proofs in the form of VCs. By processing the data entirely digitally (R 1.1), friction in the onboarding process can therefore be reduced for both the customer and the bank (Experts D, H, and I). The need for face-to-face verification, manual data processing, and repeated KYC processes can be eliminated through the use of re-usable VCs combined with revocation registries on the blockchain, thus saving costs for manual and repeated process steps (R 1.2). Expert H also emphasized that updates and periodic confirmation that customers need to provide to banks regarding their data can be significantly reduced through the bilateral and secure communication channel, through which the customer can easily give VPs to the bank. In addition to the potential personnel cost savings, the possibility of authentication with a high level of assurance and the associated reduction of risks can also avoid high penalties for non-compliance with due diligence regulations and standards. However, if there is not yet an existing ecosystem of official identity-related documents, this is only true for the *fast onboarding* process, where a prior eKYC process at another bank or official document issuer has taken place. Standards for KYC credentials can be created and stored on a public blockchain, such that they can be referred to and accepted by a range of institutions (R 1.3). This standardization can be particularly valuable when verification of unknown foreign documents can be avoided (Expert I). Nonetheless, questions regarding governance (e.g., who defines standards) remain open. An additional governance framework is therefore necessary to create clear guidelines for defining which institutions are suitable as credential issuers.

Regulatory compliance

Regarding compliance with the MLA, the interviewees did not see particular difficulties in the framework design (R 2.1). The GDPR grants the right to erasure of personal data if the reason for their processing no longer exists. While the de facto interpretation of this regulation remains unclear, it must be assumed that encrypted and hashed personal data also fall under this regulation (Expert C). Further, public DIDs and public keys could be considered as personal data under GDPR, and must therefore be deleted if customers request this (Expert E). Thus, KYC designs that use distributed ledgers to store such data cannot be implemented by banks. In our framework, natural persons only use pairwise DIDs and exchange information bilaterally without writing it to a distributed ledger (Experts E, F, and G). Further GDPR requirements, such as data minimization, are also naturally addressed through VCs' selective disclosure capabilities. The fundamental objectives of our eKYC process are therefore aligned with those of the GDPR (R 2.2). However, a detailed legal assessment remains an avenue for future research.

To effectively use the system, it must also comply with eIDAS regulation (Expert F). The experts noted that they do not see a conflict between eIDAS and the SSI-based eKYC process (R 2.3), and supported the idea of combining the SSI concept and the eIDAS infrastructure (Experts D, E, F, G, and I). Expert G stressed that “these regulations are drivers that will help to adopt SSI, because SSI is an ideal way to implement them.” The EU has started building the *eIDAS bridge*, which seeks to make the legally qualified signatures from eIDAS accessible for the VC standard; nonetheless, this implementation has not yet been completed.

Decentralization

Our framework for eKYC stores identity-related data in the customer's digital wallet, i.e., on a mobile phone or laptop. Besides the need for banks to store customer information for a certain period – owing to regulatory compliance, rather than for technical reasons – central storage is therefore unnecessary (R 3.1). User agents, whose role is discussed in R 6.3, could be considered for centralized storage. However, these only store data encrypted under user-managed keys. Further, owing to the heavy standardization associated with SSI, it is unlikely that user agents hosted by third parties will encounter the same network effects that have led to centralization for traditional identity providers in federated systems. Thus, the framework counteracts data silos that are highly attractive to hackers, since one can no longer capture many data sets at once (Experts E and F).

The proposed framework also induces no new central parties to the KYC process (R 3.2). Through the use of blockchain, no single entity controls the infrastructure that is involved in checking credential schemas or revocation registries. Expert F mentioned the banks' position of trust toward their customers, and therefore considered the banks to be very suitable providers of cloud agents and wallets. Further, most of the experts support the idea of using banks as potential service providers to backup facilities (Experts E, F, and G).

Trust

In our framework, VCs form the basis of KYC documents. The combination of VCs, as an evolution of digital certificates with additional capabilities such as selective disclosure and privacy-preserving revocation mechanisms based on a blockchain, yield a natural digital equivalent of physical KYC documents that customers can fully control and take to other banks. VCs' integrity can be tested by checking the digital signatures' validity, whereby the signing keys of issuing institutions (such as other banks) are publicly available on a blockchain. It is not possible to create fake credentials, because these are not valid without a credential issuer's signature. Further, ownership of credentials can be cryptographically proven [60], and binding multiple credentials via a strongly correlating attribute such as the holder's name or biometric properties, or cryptographically through secure hardware, makes credential sharing or selling difficult and unattractive (R 4.3). From the perspective of both banks and regulators, fully digital verification provides a significant advantage over analog documents, since data accuracy is improved and manual errors during data processing can be ruled out (R 4.1) (Experts D, E, and I).

The use of the blockchain infrastructure for storing information on credential issuers (e.g., other banks or government institutions) and revocation registries for VCs provides an infrastructure that allows a bank to verify VCs issued by other banks. Nonetheless, governance mechanisms regarding the legal acceptance of such VCs and other aspects of interbank collaboration required for (R 4.2) still leave some questions open (Experts D, F, G, and I). Such a framework is necessary to clarify which credentials the banks accept and whom they accept as a credential issuer.

Privacy

In our framework, users can store and manage their identity data independently, without having to rely on a distinguished third party. Communication is designed to be only bilateral between a credential owner and verifier, and only requires occasional, potentially anonymized read queries to a random node on a public blockchain to update schemas, issuers' signing keys, and revocation registries. This architecture prevents third parties from surreptitiously gaining insights into users' comprehensive data, as is the case with federated identity providers (R 5.1). This is also desirable from a scalability perspective (Expert G). As a result, users can have different digital identities in different contexts and only need to disclose the data required for a specific situation (Experts E and F) in accordance with the *need to know* principle.

In the SSI-based approach, customers have complete control over the data in their wallets, and customers can decide for themselves whom they wish to share data with (Experts D, E, and F). In this context, the experts also mentioned the possibility of selective disclosure through ZKP. The fact that customers no longer have to show all their personal details, but only the relevant data, helps to protect customer privacy through data minimization (R 5.2) (Experts E and F).

The experts also emphasized that a correlation of data is still possible in the absence of public identifiers on the basis of the available rich data sets of banks and other organizations. However, SSI's goal is not anonymity, as is sometimes suggested, but rather the best possible extent of privacy in each scenario. Since KYC procedures seek to build trust, a large amount of personal information must be revealed. In this context, it is important to note that researchers such as Lootsma [45] have emphasized the possibility of harnessing additional potential through KYC data by, for instance, connecting them to transactional data. While Lootsma [45] have even raised the question of resulting conflicts with customer privacy, such efforts may also lead to an inherent conflict with SSI principles. Nonetheless, once personally identifiable information has been received in plain text through a bank, it cannot be hindered in connecting it to other data, also in our approach.

User experience

The SSI-based eKYC process has the potential to vastly improve the user experience of customers. Much of the current friction, such as entering personal data in an online form, the need to visit a bank, or the need to make a video call with a bank employee for identification process, has been eliminated. Instead, the onboarding process can be carried out on the user's smartphone with just a few steps, for instance by scanning QR codes and accepting invitation links and proof requests through simple interfaces (R 6.1). Because the framework builds on generic and open standards, for which many reference implementations for mobile phones and computer operating systems are available, different user interfaces are realizable (R 6.2). Further, customers have a permanent overview of whom they shared data with (Expert E). A potential problem lies with the customer's full responsibility for data storage and administration (Experts D, E, and G) (R 6.3). Customers must develop an awareness of this so that they realize their responsibility and take appropriate backup and recovery measures to mitigate the consequences of device loss or theft (Experts E and G).

7. Discussion

As indicated in Section 4.2, our framework can greatly improve KYC processes. In particular, efficiency, trust, and privacy seem to benefit from the blockchain-based SSI architecture. However, many of the improvements do not specifically relate to the KYC case. The trust relationship illustrated in Figure 2 between a holder of ID attributes, an issuer of documents confirming these attributes, and a verifier is present in multiple domains. Thus, our architecture and its related processes reveal insights into the general design of artifacts in the nascent field of blockchain-based SSI, which according to the interviews with the SSI experts may translate to many other areas, where the fear of a centralized service provider has so far prevented a more efficient cross-organizational identity management. To elevate our IT artifact for further theoretical discussion, we derived three DPs that abstract our findings and that seek to guide future research and practice in blockchain-based SSI [29]. We analyzed codes from the interviews related to our architecture's technical building blocks (e.g., (distributed) ledger, blockchain, VC, or storage) to identify commonly proposed design patterns and their justification, and we arrived at three generic principles.

Design principle 1: Utilize blockchain only for public data

Our research suggests that the absence of a centralized platform operator in the eKYC process can enable cooperation between banks. The banks do not have to fear that other banks or even a central eKYC utility will receive valuable customer data, which could put them in a disadvantageous position or create new dependencies and lock-in effects. In this context, a distributed ledger is well suited to transparently display public information. On the other hand, owing to their inherent properties – such as transparency, redundancy, and tamper-resistance – blockchains are not suitable for storing personal data [86], even in encrypted form [17,26]. The academic literature often states that credential hashes and peer DIDs also need to be stored on a distributed ledger [53], and initial frameworks for KYC based on SSI [71] have used this approach. However, from a technical perspective and according to the experts, this has no apparent advantages and only carries performance challenges and regulatory risks: Trust in the interaction with a DID is established through a VP, and VCs' tamper resistance is established via the issuer's digital signatures, which need to be trusted anyway. This renders on-chain hashes unnecessary [78]. Further, it must be assumed that legal persons' DIDs fall under the GDPR [83], and for the aforementioned reasons, they should not be stored on a distributed ledger. Thus, distributed ledgers should only be used in a manner comparable to a public key infrastructure (PKI) for VC issuers (Experts E and F) and not for private persons' DIDs and VCs (Experts E, F, and G). By taking most communication off-chain, as Expert G mentioned, the proposed architecture and SSI could help many blockchain use cases to comply with regulation such as the GDPR or eIDAS and could resolve privacy issues (Experts C, D, E, F and G). Regarding performance, the Hyperledger Indy blockchains on which many SSI systems rely can handle only a limited number of write transactions [70]; thus, one should design interactions between stakeholders in a blockchain-based SSI environment bilaterally if possible, and one should read from a blockchain rather than write to it so as to avoid scalability issues.

To abstract and generalize this observation, we propose that by using SSI in processes that require proofs about the possession of certain attributes, organizations should repeatedly request and verify these attributes through bilateral communication channels, instead of storing the required data centrally. As a side effect, this can also help keep data up to date.

Design principle 2: Anticipate an ecosystem of various ledgers

Our initially designed framework was built on the assumption that financial institutions share a single distributed ledger to create and manage digital identities for eKYC. Employing a shared ledger facilitates interoperability on a technical level and concerning governance. However, recent developments in SSI practice [41] and our interview findings indicate that it is more likely that various distributed ledgers for SSI will exist (Experts F and G), similar to the considerable number of today's certificate authorities. Thus, it is important to account for this circumstance and to design blockchain-based SSI solutions for various distributed ledgers to achieve interoperability. This can be achieved through adherence to industry standards, which are currently being developed by organizations such as the W3C [73], as well as by using technical components for interoperability and trust. Universal resolvers – i.e., identifier resolvers working for a multitude of identifiers such as DIDs on different blockchains and maybe also centralized databases (e.g., provided by certificate authorities) – may play an important role in this regard and may also increase trust (Experts D, E, F, and G). While interoperability is technically achievable without major challenges, the existence of various distributed ledgers may induce governance-level challenges that must also be designed through cross-ledger governance (Expert G).

Design principle 3: Enable decentralization at the edge

During the creation of the SSI-based KYC framework, we encountered some challenges regarding SSI-based identity management's user-friendliness. Although managing all identity-relevant data through a single app can boost the straightforward and user-friendly management of identity documents and increase authenticity through hardware-binding or credential-linking, self-managing leads to multiple challenges. For instance, users need support if their devices are lost or stolen. The status quo of central systems must be broken down somewhat here [83], and users must develop an awareness of their responsibility for their data and must learn to store them accordingly (Experts E and F). On the other hand, one must also find the right balance between decentralized and central solutions [21]. An example can be the use of cloud storage and cloud agents, which can add value concerning recovery, availability, and security if the agent specializes in this service. On the other hand, these cloud solutions contradict SSI's basic idea of avoiding as many third parties as possible, particularly honey-pots of data, for privacy and security reasons. However, as long as the data are encrypted and the cloud providers cannot access the data, Expert C sees cloud storage a both a viable and an essential element for enabling good user experience. To think decentralization to an end and support the autonomy of end users in blockchain-based SSI applications, SSI-based architectures must ensure that users can store their VCs on an infrastructure of their choice.

Crossing the chasm: How to bring blockchain-based SSI into practice in KYC and beyond

While design artifacts, as the outcome of the DSR process, should have practical impact [7], bringing the artifacts into practice requires suitable approaches. One key topic mentioned by multiple experts – but that does not relate to technical terms and is therefore not a DP that we can derive based on our codes related to technical building blocks – is the adoption of the SSI-based eKYC. These experts regarded the general adoption of SSI technology in the public and private sectors as a major driving force of the practical implementation of our eKYC solution. Expert D called this a *chicken and egg* problem: Since the technology is still very new, there are few credential providers, so the utility for a user is very low; on the other hand, as long as there are only a few users, there is also no major incentive for organizations to act as credential issuers. The more credentials users have, the better such a system can be used (Expert D). Network effects can help bridge the gap between early adopters and the widespread use of the technology [49]. In particular, banks can contribute to this by offering the technology in the *isolated* KYC use case, issuing VCs to contribute to its spread in the mass market. The more that banks accept these credentials, the more attractive the system becomes to customers. In turn, higher usage by customers leads to more incentive for other organizations to accept VCs. As mentioned, the cooperation of the banks and the creation of shared standards are crucial if this adoption is to become possible.

Because users have full control over their data, SSI must address the GDPR's general requirements, such as privacy by design, portability, the right to erasure, transparency, purpose limitation, data minimization, accuracy, storage limitation, and information integrity. Users can get a permanent overview of whom they shared what data with (Expert E), and these records can help to better implement the right to erasure. In turn, this may lead to a better adoption of the technology. In fact, the GDPR's strict requirements, which were often criticized for impeding blockchain-related innovation in Europe, may ultimately have turned out to boost innovation, so blockchain's benefits for interoperability can be used for the purposes highlighted in DP 1 while avoiding its well-known privacy- and scalability-related challenges.

The interplay between SSI and regulation uncovers many other interesting dimensions. For instance, our interviewees suggested that eIDAS regulation can facilitate SSI while SSI can help make the eIDAS infrastructure, which so far has been used only moderately, more practicable and valuated [77]. On the other hand, we saw that SSI technically allows for even more privacy than what is required by regulation. However, the MLA requires that banks store customer data for five years, creating tension between data protection regulation and the objectives of user control and the prevention of data silos. Thus, SSI may even lead to new discussions on where to set the sweet spot between market integrity and privacy.

8. Conclusion

In this article, we sought to build a framework to improve on the current shortcomings in the KYC process through an end-to-end digital process that leverages blockchain-based SSI. Research on SSI is still in its infancy, and little has been published on the design of applications for SSI. Soltani et al. [71] were the first to explore this topic in the context of KYC, covering the onboarding process and technically evaluating their solution. Building on this valuable work, we extended the scope and emphasized banks' requirements. We used a DSR approach based on Peffers et al. [57], designing and evaluating a framework for KYC processes built on blockchain-based SSI, including a generic architecture and process design. Since we face a low solution maturity in the innovative field of blockchain and SSI, and high application domain maturity in the domain of KYC, we provided an *improvement* in the context of DSR [29]. Our evaluation suggests that our design can significantly contribute to a more efficient KYC process that also addresses the other requirements of stakeholders. Thus, we are confident that we have accomplished our research objective.

Besides the conceptualized and evaluated architecture and set of processes [29] for the KYC case, we made three primary contributions to the academic body of knowledge. First, our examination revealed the challenges of using DLT for the exchange of personal data generally and particularly for digital identity management systems. We also showed how these problems can be solved by using SSI on top of the blockchain layer, thereby leveraging the advantages typically associated with blockchain technology while avoiding its well-known issues with scalability and privacy. Second, we revealed the implications of designing SSI-based solutions built on blockchain in the context of KYC by deriving three DPs, which allowed us to elevate our IT artifact for more abstract and generalizable theoretical discussion [29]. Third, we offered suggestions for relevant future research on blockchain and SSI, enabling researchers to base their work on our results and thus generate additional knowledge [82].

DSR should also inform practice to advance a specific domain through IT [29]. Our conception and evaluation of the SSI-based KYC framework will provide practitioners with valuable insights regarding design choices, DLT's role, the intricacies of regulation, and related challenges and opportunities for banks and customers. Our results indicate that SSI-based eKYC processes can reduce cost and time expenditures and contribute to better user experiences and increased security during the KYC process. We demonstrated how the use of SSI can positively impact the different onboarding processes and their interplays with an existing SSI ecosystem. However, we illustrated that there are further conceptual challenges to be solved before SSI is used in real systems and settings, especially regarding the necessary governance frameworks and a more detailed regulatory analysis. While our research suggests synergies between SSI and regulation, challenges remain, especially to establish a general SSI-based ecosystem and to make SSI as user-friendly as possible without sacrificing privacy and security.

Like most research, our study has limitations. Our framework has not yet been used in practice and therefore lacks an evaluation in a real-world setting. However, by applying a rigorous research design and obtaining practitioner feedback, we sought to address this shortcoming. Further, although we described the necessary and central elements of an additionally required governance framework, its concrete implementation remains open. In particular, details regarding the cooperation of banks in the KYC process, the creation of shared standards, and the responsible parties for operating the blockchain in the case of using a permissioned network must be clarified. This opens various promising avenues for future research. In particular, the KYC process is often relevant not only in the banking environment but also in other domains such as insurance, and further objectives may be necessary to address this aspect. Further, although our interviews with experts confirmed Moyano and Ross's [51] findings that there is sufficient trust between banks that collaboration on eKYC is possible despite competition on a suitable IT infrastructure, this only represents the perspective of practitioners and researchers in central and northern European countries. Nonetheless, there is also a promising development that may make SSI-based KYC and our findings considerably more far-reaching and applicable in contexts in which this trust is missing: In emerging SSI ecosystems in North America and Europe, governments are starting to explore the impacts of issuing certificates such as driver's licenses and ID cards in the form of VCs that can be leveraged by the public and private sector. Besides increasing the efficiency of digital identification and authentication, a digital ID is expected to contribute to substantial reductions in financial crime [25]. In this context, an experimental clause was recently adopted in Germany's parliament that explicitly allows banks to perform KYC based on an ID card in the form of a VC [5,19].

We are confident that we have derived guidelines and DPs that generalize to these promising developments and to other sectors, and that also provide guidance on how to make blockchain-based SSI compatible with the needs of businesses and regulatory restrictions. More efficient, reusable processes, specifically relating to identity management, are needed in both business and the public sector, but the risks associated with central providers have so far prevented general services from providing these capabilities. Blockchain-based SSI can address the need for a general service for data that is non-competitive, since the same data are needed and used by all, and yet comply with customer data privacy expectations and regulations. On the other hand, our DP of minimal involvement of the blockchain, specifically not storing natural persons' DIDs or even VCs on a blockchain, translates to applications of SSI generally, and we are eager to see more use cases being built on this technology stack. Given current efforts by the Verifiable Organizations Network in Canada, the ambitious goal of having 10 different pilots that leverage blockchain-based SSI by the end of 2021 in Germany, and several SSI smartphone wallets that are already available, we are confident that this time, the promises of blockchain-technology to revolutionize digital ID management can be fulfilled, although blockchain's role will be much more restricted than what early research suggested.

CRediT authorship contribution statement

Vincent Schlatt: Conceptualization, Funding acquisition, Formal analysis, Writing – original draft, Writing – review & editing. **Johannes Sedlmeir:** Funding acquisition, Formal analysis, Writing – original draft, Writing – review & editing. **Simon Feulner:** Conceptualization, Funding acquisition, Formal analysis, Writing – original draft, Writing – review & editing. **Nils Urbach:** Conceptualization, Formal analysis, Writing – review & editing.

References

[1] C. Allen, The path to self-sovereign identity, 2016, http://www.lifewithalacrity.com/previous/2016/04/the-path-to-self-sovereign-identity.html.

[2] R. Arasa, L. Ottichilo, Determinants of know your customer (KYC) compliance among commercial banks in Kenya, Journal of Economics and Behavioral Studies 2 (2015) 162–175.

[3] D.W. Arner, J.N. Barberis, R.P. Buckley, The emergence of regtech 2.0: From know your customer to know your data, 2016, https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3044280_code722134.pdf?abstractid=3044280mirid=1.

[4] D.W. Arner, D.A. Zetzsche, R.P. Buckley, J.N. Barberis, The identity challenge in finance: From analogue identity to digitized identification to digital KYC utilities, European Business Organization Law Review 20 (1) (2019) 55–80.

[5] Association of German Banks. digital identities – steps on the path to an ID ecosystem, 2021, https://en.bankenverband.de/newsroom/comments/digital-identities-steps-path-id-ecosystem/#2.

[6] O. Avellaneda, A. Bachmann, A. Barbir, J. Brenan, P. Dingle, K.H. Duffy, E. Maler, D. Reed, M. Sporny, Decentralized identity: Where did it come from and where is it going? IEEE Communications Standards Magazine 3 (4) (2019) 10–13.

[7] R. Baskerville, A. Baiyere, S. Gregor, A. Hevner, M. Rossi, Design science research contributions: Finding a balance between artifact and theory, Journal of the Association for Information Systems 19 (5) (2018) 3.

[8] R. Beck, S. Weber, R.W. Gregory, Theory-generating design science research, Information Systems Frontiers 15 (4) (2013) 637–651.

[9] A. Biryukov, D. Khovratovich, S. Tikhomirov, Privacy-preserving KYC on Ethereum. 1st Blockchain Workshop, ERCIM, 2018.

[10] B.-J. Butijn, D.A. Tamburri, W.J.v.d. Heuvel, Blockchains: A systematic multivocal literature review, ACM Computing Surveys 53 (3) (2020) 1–37.

[11] J. Camenisch, A. Lysyanskaya, An efficient system for non-transferable anonymous credentials with optional anonymity revocation. International Conference on the Theory and Applications of Cryptographic Techniques, Springer, 2001, pp. 93–118.

[12] F. Casino, T.K. Dasaklis, C. Patsakis, A systematic literature review of blockchain-based applications: Current status, classification and open issues, Telematics and Informatics 36 (2019) 55–81.

[13] K. Charmaz, Constructing grounded theory: A practical guide through qualitative analysis, Sage, 2006.

[14] R. Christie, Setting a standard path forward for KYC, Journal of Financial Transformation 47 (2018) 155–164.

[15] S. Clauß, M. Köhntopp, Identity management and its support of multilateral security, Computer Networks 37 (2) (2001) 205–219.

[16] J. Corbin, A. Strauss, Basics of qualitative research: Techniques and procedures for developing grounded theory, Sage, 2008.

[17] COVID-19 Credential Initiative. Decentralised Identity Architecture and Regulatory Compliance, 2020, https://docs.google.com/document/d/164S-nNnMHlPZex5lKenuGRhavvt8qeKIdTPHFAD5b_c/edit.

[18] M. Davie, D. Gisolfi, D. Hardman, J. Jordan, D. O'Donnell, D. Reed, The trust over IP stack, IEEE Communications Standards Magazine 3 (4) (2019) 46–51.

[19] Deutscher Bundestag. drucksache 19/30443, 2021, https://dserver.bundestag.de/btd/19/304/1930443.pdf.

[20] R. Dhamija, L. Dusseault, The seven flaws of identity management: Usability and security challenges, IEEE Security & Privacy 6 (2) (2008) 24–29.

[21] P. Dunphy, F.A. Petitcolas, A first look at identity management schemes on the blockchain, IEEE Security & Privacy 16 (4) (2018) 20–29.

[22] T.E. Maliki, J.M. Seigneur, A survey of user-centric identity management technologies, The International Conference on Emerging Security Information, Systems, and Technologies (2007) 12–17.

[23] European Commission. eIDAS supported self-sovereign identity, 2019, https://ec.europa.eu/futurium/en/system/files/ged/eidas_supported_ssi_may_2019_0.pdf.

[24] FATF, FATF 40 recommendations, 2004, https://www.fatf-gafi.org/media/fatf/documents/FATF%20Standards%20-%2040%20Recommendations%20rc.pdf.

[25] Financial Times. 81 % of financial services firms agree digital ID would improve financial crime prevention, 2021, https://thefintechtimes.com/81-of-financial-services-firms-agree-digital-id-would-improve-financial-crime-prevention/.

[26] M. Finck, Blockchains and data protection in the european union, European Data Protection Law Review 4 (2018) 17–35.

[27] G. Fridgen, S. Radszuwill, N. Urbach, L. Utz, Cross-organizational workflow management using blockchain technology – towards applicability, auditability, and automation, Proceedings of the 51st Hawaii International Conference on System Sciences (2018) 3507–3516.AIS

[28] F. Glaser, Pervasive decentralisation of digital infrastructures: A framework for blockchain enabled system and use case analysis. Proceedings of the 50th Hawaii International Conference on System Sciences, AIS, 2017, pp. 1543–1552.

[29] S. Gregor, A.R. Hevner, Positioning and presenting design science research for maximum impact, MIS Quarterly 37 (2) (2013) 337–355.

[30] T. Guggenberger, A. Schweizer, N. Urbach, Improving interorganizational information sharing for vendor managed inventory: Toward a decentralized information hub using blockchain technology, IEEE Transactions on Engineering Management (2020).

[31] D. Hardman, No paradox here: ZKPs deliver savvy trust, 2020, https://www.evernym.com/blog/no-paradox-here-zkps-deliver-savvy-trust/.

[32] D. Hardman, L. Harchandani, A. Othman, J. Callahan, Using biometrics to fight credential fraud, IEEE Communications Standards Magazine 3 (4) (2019) 39–45.

[33] A. Hevner, S. Gregor, Envisioning entrepreneurship and digital innovation through a design science research lens: A matrix approach, Information & Management (2020) 103350.

[34] A. Hevner, S.T. March, J. Park, S. Ram, et al., Design science research in information systems, MIS Quarterly 28 (1) (2004) 75–105.

[35] P.L. Hughes, The 'need to know' principle of computer security, Computer Law & Security Review 3 (5) (1988) 29–30.

[36] B. Jessel, K. Lowmaster, N. Hughes, et al., Digital identity: The foundation for trusted transactions in financial services, Journal of Financial Transformation 47 (2018) 143–150.

[37] D. Jones, S. Gregor, The anatomy of a design theory, Journal of the Association for Information Systems 8 (5) (2007) 312–335.

[38] H. Kallio, A.-M. Pietilä, M. Johnson, M. Kangasniemi, Systematic methodological review: Developing a framework for a qualitative semi-structured interview guide, Journal of Advanced Nursing 72 (12) (2016) 2954–2965.

[39] S. Kokolakis, Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon, Computers & Security 64 (2017) 122–134.

[40] J. Kolb, M. AbdelBaky, R.H. Katz, D.E. Culler, Core concepts, challenges, and future directions in blockchain: A centralized tutorial, ACM Computing Surveys 53 (1) (2020) 1–39.

[41] M. Kuperberg, Blockchain-based identity management: A survey from the enterprise and ecosystem perspective, IEEE Transactions on Engineering Management 67 (4) (2019) 1008–1027.

[42] Ledger Insights. Self-sovereign identity successfully trialed for KYC in UK regulatory sandbox, 2020, https://www.ledgerinsights.com/self-sovereign-identity-successfully-trialed-for-kyc-in-uk-regulatory-sandbox/.

[43] S.Y. Lim, P.T. Fotsing, A. Almasri, O. Musa, M.L.M. Kiah, T.F. Ang, R. Ismail, 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.

[44] Y. Liu, Q. Lu, H.-Y. Paik, X. Xu, S. Chen, L. Zhu, Design pattern as a service for blockchain-based self-sovereign identity, IEEE Software 37 (5) (2020) 30–36.

[45] Y. Lootsma, Blockchain as the newest regtech application – the opportunity to reduce the burden of KYC for financial institutions, Banking & Financial Services Policy Report 36 (8) (2017) 16–21.

[46] E. Maler, D. Reed, The Venn of identity: Options and issues in federated identity management, IEEE Security & Privacy 6 (2) (2008) 16–23.

[47] S.T. March, G.F. Smith, Design and natural science research on information technology, Decision Support Systems 15 (4) (1995) 251–266.

[48] J.H. Moor, Towards a theory of privacy in the information age, ACM Sigcas Computers and Society 27 (3) (1997) 27–32.

[49] G.A. Moore, R. McKenna, Crossing the chasm, Harper Business Essentials, 1999.

[50] J.M. Morse, Strategies for sampling, Qualitative nursing research: A contemporary dialogue (1991) 127–145.

[51] J.P. Moyano, O. Ross, KYC optimization using distributed ledger technology, Business & Information Systems Engineering 59 (6) (2017) 411–423.

[52] N. Mugarura, Customer due diligence (CDD) mandate and the propensity of its application as a global AML paradigm, Journal of Money Laundering Control 17 (1) (2014) 75–96.

[53] A. Mühle, A. Grüner, T. Gayvoronskaya, C. Meinel, A survey on essential components of a self-sovereign identity, Computer Science Review 30 (2018) 80–86.

[54] M.D. Myers, M. Newman, The qualitative interview in IS research: Examining the craft, Information and organization 17 (1) (2007) 2–26.

[55] R. Norvill, M. Steichen, W.M. Shbair, R. State, Blockchain for the simplification and automation of KYC result sharing. International Conference on Blockchain and Cryptocurrency, IEEE, 2019, pp. 9–10.

[56] N.K. Ostern, J. Riedel, Know-your-customer (KYC) requirements for initial coin offerings, Business & Information Systems Engineering (2020), https://doi.org/10.1007/s12599-020-00677-6.ISSN 1867-0202

[57] K. Peffers, T. Tuunanen, M.A. Rothenberger, S. Chatterjee, A design science research methodology for information systems research, Journal of Management Information Systems 24 (3) (2007) 45–77.

[58] L. Perlman, N. Gurung, Focus note: The use of eKYC for customer identity and verification and AML, 2019, https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3370665_code505438.pdf?abstractid=3370665mirid=1.

[59] A. Rajput, K. Gopinath, Towards a more secure aadhaar. International Conference on Information Systems Security, Springer, 2017, pp. 283–300.

[60] K. Rannenberg, J. Camenisch, A. Sabouri, Attribute-based credentials for trust. Identity in the Information Society, Springer, 2015.

[61] D. Reed, J. Law, D. Hardman, M. Lodder, DKMS (decentralized key management system) design and architecture v3, 2018, https://github.com/hyperledger/indy-sdk/blob/677a0439487a1b7ce64c2e62671ed3e0079cc11f/doc/design/005-dkms/DKMS20Design20and20Architecture20V3.md.

[62] D. Reed, M. Sporny, D. Longley, C. Allen, R. Grant, M. Sabadello, Decentralized identifiers (DIDs) v1.0, 2020, https://w3c.github.io/did-core.

[63] K. Reinecke, A. Bernstein, Knowing what a user likes: A design science approach to interfaces that automatically adapt to culture, MIS Quarterly (2013) 427–453.

[64] A. Rieger, J. Lockl, N. Urbach, F. Guggenmos, G. Fridgen, Building a blockchain application that complies with the EU general data protection regulation, MIS Quarterly Executive 18 (4) (2019).

[65] M. Rossi, C. Mueller-Bloch, J.B. Thatcher, R. Beck, Blockchain research in information systems: Current trends and an inclusive future research agenda, Journal of the Association for Information Systems 20 (9) (2019) 14.

[66] P.J. Ruce, Anti-money laundering: The challenges of know your customer legislation for private bankers and the hidden benefits for relationship management (the bright side of knowing your customer), The Banking Law Journal 128 (6) (2011) 548–564.

[67] J. Saldaña, The coding manual for qualitative researchers, Sage, 2015.

[68] A. Schweizer, V. Schlatt, N. Urbach, G. Fridgen, Unchaining social businesses – blockchain as the basic technology of a crowdlending platform, 37th International Conference on Information Systems (2017).

[69] J. Sedlmeir, R. Smethurst, A. Rieger, G. Fridgen, Digital identities and verifiable credentials, Business & Information Systems Engineering (2021).

[70] J. Sedlmeir, P. Ross, A. Luckow, J. Lockl, D. Miehle, G. Fridgen, The DLPS: A framework for benchmarking blockchains. Proceedings of the 54th Hawaii International Conference in System Sciences, IEEE, Wailea, Maui, Hawaii, USA, 2021, pp. 6855–6864.

[71] R. Soltani, U.T. Nguyen, A. An, A new approach to client onboarding using self-sovereign identity and distributed ledger. International Conference on Internet of Things and Green Computing and Communications and Cyber, Physical and Social Computing and Smart Data, IEEE, 2018, pp. 1129–1136.

[72] C. Sonnenberg, J.V. Brocke, Evaluations in the science of the artificial – Reconsidering the build-evaluate pattern in design science research. International Conference on Design Science Research in Information Systems, Springer, 2012, pp. 381–397.

[73] M. Sporny, D. Longley, D. Chadwick, Verifiable credentials data model 1.0, 2019, https://www.w3.org/TR/vc-data-model/.

[74] D. Swinhoe, The 15 biggest data breaches of the 21st century, 2020, https://www.csoonline.com/article/2130877/the-biggest-data-breaches-of-the-21st-century.html.

[75] The Economist. Indian business prepares to tap into Aadhaar, a state-owned fingerprint-identification system, 2016, https://www.economist.com/business/2016/12/24/indian-business-prepares-to-tap-into-aadhaar-a-state-owned-fingerprint-identification-system.

[76] Thomson Reuters. know your customer surveys reveal escalating costs and complexity, 2016, https://www.thomsonreuters.com/en/press-releases/2016/may/thomson-reuters-2016-know-your-customer-surveys.html.

[77] K.T.T. Lyons, L. Courcelas, Blockchain and digital identity, 2019, https://www.eublockchainforum.eu/sites/default/files/report_identity_v0.9.4.pdf.

[78] K.C. Toth, A. Anderson-Priddy, Self-sovereign digital identity: A paradigm shift for identity, IEEE Security & Privacy 17 (3) (2019) 17–27.

[79] Trust over IP Foundation. Introducing the Trust over IP Foundation, 2020, https://trustoverip.org/wp-content/uploads/sites/98/2020/05/toip_introduction_050520.pdf.

[80] J. Venable, J. Pries-Heje, R. Baskerville, A comprehensive framework for evaluation in design science research. International Conference on Design Science Research in Information Systems, Springer, 2012, pp. 423–438.

[81] J. Venable, J. Pries-Heje, R. Baskerville, FEDS: A framework for evaluation in design science research, European Journal of Information Systems 25 (1) (2016) 77–89.

[82] J.V. Brocke, R. Winter, A. Hevner, A. Maedche, Special issue editorial – accumulation and evolution of design knowledge in design science research: A journey through time and space, Journal of the Association for Information Systems 21 (3) (2020) 9.

[83] K. Wagner, B. Némethi, E. Renieris, P. Lang, E. Brunet, E. Holst, Self-sovereign identity: A position paper on blockchain enabled identity and the road ahead, 2018, https://jolocom.io/wp-content/uploads/2018/10/Self-sovereign-Identity-_-Blockchain-Bundesverband-2018.pdf.

[84] L. Zavolokina, R. Ziolkowski, I. Bauer, G. Schwabe, Management, governance and value creation in a blockchain consortium, MIS Quarterly Executive 19 (1) (2020) 1–17.

[85] D.A. Zetzsche, R.P. Buckley, D.W. Arner, Digital ID and AML/CDD/KYC utilities for financial inclusion, integrity and competition, Journal of Financial Transformation (2018) 133–142.

[86] R. Zhang, R. Xue, L. Liu, Security and privacy on blockchain, ACM Comput. Surv. 52 (3) (2019).

Vincent Schlatt (vincent.schlatt@fit.fraunhofer.de) is a researcher at the Project Group Business & Information Systems Engineering of the Fraunhofer FIT and currently pursuing his PhD in Information Systems at the FIM Research Center, University of Bayreuth. His research interests focus on decentralized technologies, such as blockchain and self-sovereign identity, and design science research methods. He received his M.Sc. in Information Technology.

Johannes Sedlmeir (johannes.sedlmeir@fim-rc.de) is a researcher at the Project Group Business & Information Systems Engineering of the Fraunhofer FIT and currently pursuing his PhD in Information Systems at the FIM Research Center, University of Bayreuth. In his research, Mr. Sedlmeir works on the energy consumption and performance benchmarking of different blockchains, decentralized digital identities, and the application of cryptographic methods such as Zero-Knowledge Proofs for scalability and privacy on blockchains. He received his M.Sc. in Theoretical and Mathematical Physics.

Simon Feulner (simon.feulner@fim-rc.de) is a research associate at the Project Group Business & Information Systems Engineering of the Fraunhofer FIT and currently pursuing his PhD in Information Systems at the FIM Research Center, University of Bayreuth. In his research, Mr. Feulner focuses on the impact of distributed technologies, such as blockchain and self-sovereign identity, the concept of machine economy, and design science research. He received his M.Sc. in Business Administration and Information Systems.

Dr. Nils Urbach (nils.urbach@fim-rc.de) is Professor of Information Systems, Digital Business and Mobility at the Frankfurt University of Applied Sciences, Germany. He is also the Deputy Director of the Research Center Finance & Information Management (FIM) and the Project Group Business & Information Systems Engineering of Fraunhofer FIT. His works have been published in several international journals such as the Journal of Information Technology, Journal of Strategic Information Systems, and Journal of Information Technology Theory and Application as well as in the proceedings of key international conferences.