POLITECNICO
MILANO 1863

School of Industrial and Information Engineering

Master of Science in Management Engineering

SELF-SOVEREIGN IDENTITY: STUDY OF THE MODEL AND CENSUS OF INTERNATIONAL PROJECTS

Supervisor:

Luca Gastaldi

Author:

Francesco Vitali, 969534

Academic Year 2021/2022

Abstract

In recent years, one of the most widely used terms in the landscape of identity management is Self-Sovereign Identity. It is the latest evolution of identity management models. Self-Sovereign Identity is a decentralized identity model that allows users to have full control of their digital identity. This type of identity responds to the need for decentralization of identification systems and for restoring control of one's identity to people. Although fragmented, there are several theoretical academic writings on the Self-Sovereign Identity model. On the contrary, at the application level, scientific articles are lacking. The objective of this work is to address this gap, in order to understand how the landscape of Self-Sovereign Identity type systems is configured at an international level. For this purpose, a census was carried out, which made it possible to identify 51 cases of Self-sovereign Identity. The 51 projects were identified through research on multiple sources. Furthermore, for each case identified, the information on the variables considered interesting was integrated using secondary sources. The variables considered relevant were identified following an in-depth review of the existing literature. The results of the subsequent analyses are discussed and interpreted in the light of the existing literature on Self-Sovereign Identity. In this way it was possible to confirm some theoretical findings and intercept the main application trends of Self-Sovereign Identity.

Keywords: Self-Sovereign Identity, Self Sovereign Identity, Digital Identity, Decentralized Identity, Identification, Authentication, Verification, Decentralized Identifiers, Verifiable Credentials, Wallet, Key Cryptography, ZKP, Blockchain

Abstract in italiano

Negli ultimi anni, uno dei termini più utilizzati nel panorama dell'identità digitale è Self-Sovereign Identity. Si tratta dell'ultima evoluzione dei modelli di gestione delle identità. La Self-Sovereign Identity è un modello di identità decentralizzato che consente agli utenti di avere il pieno controllo della propria identità digitale. Questo tipo di identità risponde all'esigenza di decentralizzazione dei sistemi di identificazione e permette di restituire il controllo dei propri dati identificativi agli individui. Sebbene frammentate, esistono diverse pubblicazioni accademiche sul modello teorico della Self-Sovereign Identity. Al contrario, vi è una mancanza di articoli scientifici a livello applicativo. L'obiettivo di questo lavoro è affrontare questa mancanza, al fine di comprendere come si configura a livello internazionale il panorama dei sistemi di tipo Self-Sovereign Identity. A tal fine è stato effettuato un censimento che ha permesso di individuare 51 casi di Self-Sovereign Identity. I 51 progetti sono stati individuati attraverso ricerche mirate su più fonti. Inoltre, per ogni caso individuato, le informazioni sulle variabili ritenute interessanti sono state integrate utilizzando fonti secondarie. Le variabili ritenute rilevanti sono state individuate a seguito di un'approfondita revisione della letteratura esistente. I risultati delle analisi successive sono discussi e interpretati alla luce della letteratura esistente sulla Self-Sovereign Identity. In questo modo è stato possibile confermare alcuni riscontri teorici e intercettare i principali trend applicativi della Self-Sovereign Identity.

Parole chiave: Self-Sovereign Identity, Self Sovereign Identity, Identità Digitale, Identità Decentralizzata, Identificazione, Autenticazione, Verifica, Identificatori Decentralizzati, Credenziali Verificabili, Wallet, Chiavi Crittografiche, ZKP, Blockchain

Contents

Executive Summary

The evolution of the world is moving towards an increasing digitization, as well as an increasing use of the web and digital technologies. The advent of the Covid-19 pandemic has further increased this growth. Companies have had to make their people work from home in smart working, further increasing the need for digitization. Also the public sector has been affected, moving many of the services that were previously available in government offices online. In this context, having a reliable digital identity is essential and often mandatory to be able to fully exploit the advantages of digitization. It is therefore necessary to be able to certify a certain amount of information that allows to create an environment of trust even online, but at the same time it is necessary to protect the privacy and security of users, increasingly threatened in recent times. Digital identity comes with several benefits, such as the reduction of costs and times with the consequent increase in the sales of goods and services, which in turn improves employment and work productivity. From a social point of view, digital identity could favor the achievement of goal 16.9 of the Sustainable Development Goals set by the United Nations in 2015: “to provide legal identity to all by 2030”. It is estimated that in 2020 there were 1.1 billion people still without an identity. This excludes such people from society, as it prevents them, for example, from voting or opening a bank account.

When it comes to digital identity, the reference model is the centralized one. This model, in the face of some consolidated advantages, presents some problems on which more and more attention is being paid. These include problems of security, respect of the privacy rights of end users and the loss of control over one's identity by the legitimate owners. For this reason, in recent years, one of the most discussed terms in the identity management landscape is Self-Sovereign Identity. It is a decentralized identity model, which allows to overcome the limitations of the other identity systems, restituting users control over their own digital identity, while at the same time guaranteeing greater privacy and security. In addition, with the exponential growth of blockchain technology in recent years, there is optimism about the transfer of that growth to the Self-Sovereign Identity as well. Indeed, although not essential, the blockchain has several benefits and favors the practical application of this identity model. Self-Sovereign Identity is believed to have a strong effect on how we interact with each other on the internet in the future.

This work aims at examining the ecosystem of Self-Sovereign Identity, to discover the characteristics and the types of solution and service offered related to the topic. To achieve the objective, the research methodology was composed of three fundamental parts. The first is the theoretical review of the existing literature, with an in-depth analysis of documents, articles, and reports on the subject. From this first part it emerged that the existing literature presents multiple theoretical writings on the subject. However, there is a lack of documents regarding the practical applications of the model. For this reason, in the second part, a census of the existing practical cases of Self-

Sovereign Identity was carried out. The goal was to understand how the landscape of Self-Sovereign Identity systems is configured at an international level. This was possible based on what was learned during the literature review process. Indeed, the empirical framework, used to construct the census, was created precisely starting from there. At the end, the census counts 51 cases of Self-Sovereign Identity accompanied by a multitude of information on the relevant variables identified. In the last part, analyses were conducted in which the variables were analyzed individually and combined, to collect the main insights and trends of the ecosystem.

The ecosystem is unbalanced towards the European continent, with Germany being the most represented country. Most of the projects (61%) are in an experimental phase (PoC, Pilot, etc.), while only a quarter of the SSI systems surveyed are active. However, the number of active projects has been growing in recent years. The private sector is driving the model, being represented in all cases, and being present exclusively in most of them. A multitude of different **SSI protocols** (23) were identified during the research, among them Sovrin, Jolocom and the W3C Standards emerge as references in the ecosystem.

From the analysis of each of the ten **principles of the SSI model**, it emerged that some principles are respected more easily than others. Among the most critical principles are Persistence, Minimization, Portability, and Interoperability. The last two are not respected or are only partially respected in more than half of the cases.

As for the **technological aspects**, the practical analysis confirmed the strong link between the SSI model and the Blockchain, used in 88% of cases. These blockchains are mostly public (82%) and permissioned (56%). Considering, instead, the integration technologies to dominate are API & SDK used in all cases with known information. While, from the point of view of process technologies, mobile applications are used in all cases, often in combination with a wallet (79% of cases).

Moving on to the perspective of **users and their experience**, it can be said that the SSI model relies mainly on innovative methods, detaching itself from the more traditional procedures of digital identity. This is evident considering the results on access methods, which show a decline in the use of User ID and Password, in favor of biometric factors, usable in 91% of cases. The trend is also valid considering the onboarding procedures; indeed, it is possible to carry out online onboarding in more cases than those in which traditional in-person onboarding is required. Theoretically, in SSI systems any attribute could be associated with identity. However, in 54% of cases it is possible to attribute only two or fewer data types to the identity (personal documents and driving license are the most frequently integrated data), effectively delineating a limited number of possible attributes for the identity. Projects that allow at least four types of data to be attributed to identity are only 25% of the total. The data described so far are always certified or linked to a physical document.

Lastly, considering the **application areas**, the most popular projects are general purpose ones. Of all cases, 27% are of this type. Among the vertical application fields, the most common are finance, eGov, healthcare and mobility. There are projects in other areas too, but they are less numerous.

Although this study is not without limitations, its findings contribute to the creation of a new and reliable database on Self-Sovereign Identity cases, from which future research can start. However, the most important contribution of this work remains the identification of the most widespread practices and the main application trends of SSI systems, highlighting successful projects and those that have encountered difficulties. Thus, supporting greater development of the SSI ecosystems at an international level, pushing digital identity towards a new paradigm that brings the user back to the center of the system.

Introduction

The rise of the Internet and Web 2.0 in recent years has led to an evolution in the way interactions between digital entities are performed. Large areas of business have embarked on a digital transition process for their services, such as online banking, e-commerce, messaging, and travel booking. In addition, the Covid-19 pandemic has given a further incentive to the digitization of services in both the public and private sectors. At the private level, the financial sector has seen the most notable change in terms of digital transformation. In this sector it is mandatory by law to have a precise system of electronic identification and recognition of the user, for example to have access to a bank account (Arner et al., 2019). Moving to the public sector, some governments, such as Estonia, have taken action in recent years to build pervasive digital identity ecosystems. This is dictated by the desire to overcome the inefficiencies of the systems in place, towards highly integrated and interoperable digital economies (Atick, 2016). This continuous transition to the digital world has profoundly changed the behavior of companies and end users. As for businesses, a digital-oriented economy requires looking for new ways to interact with customers, a change in the ways they market their products. Businesses' relationships have not only changed between customers and products, but also with partners, suppliers, and employees. For customers, the changes are equally radical. Products and services are no longer purchased only in physical locations, but also online. Furthermore, the classic trust patterns that most people have relied on are absent or can be falsified in the digital context (Dib & Toumi, 2020).

Also the way people interact with each other has changed. Social media, like WhatsApp or Instagram, have revolutionized the way people interact and share experiences. Computers and mobile phones have become the hub of the mobile telecommunications industry, enabling the development of an interconnected society.

Specific needs and requirements in these various use cases have driven the development of different identity management systems. The ultimate consequence of this is that users are left with a large number of identities, which they often struggle to manage. For these reasons having an efficient management of digital identity and a better control of interactions is necessary to be able to cope with these changes (Dib & Toumi, 2020).

Digital Identity is a topic born about 20 years ago, finding first evidence in papers like "Digital identity", written by Camp in 2004, or "Digital identity matters", written by Allison et al. in 2005. Its importance has become increasingly clear due to the necessity of a proof of existence in the digital world and since it allows to assess and authenticate an entity on the web, without necessarily involving human operators. The currently prevailing model in the digital identity field is that of Identity as a Service (IdaaS). This model allows for interesting applications, such as single sign on, and is generally considered to be efficient. However, it is not very flexible as the data that can be

guaranteed is defined in advance and it is difficult to extend it. Furthermore, this model is centralized, an attack on an identity provider prevents users from identifying themselves. Even more so because centralized models contain a large amount of information, making them attractive targets for hackers, leading to an increase in security breaches and identity fraud. As the number of identity breaches increases, awareness of the implications associated with existing digital identity management approaches and their shortcomings increases. This is because each breach results in a significant loss of personal data and enormous costs for all parties involved, especially users. A further problem is the lack of adequate data ownership and control over digital identity data by users, as well as the absence of adequate digital identity for over a billion people worldwide. This has a negative impact on users' privacy rights and access to services. Additionally, password-based authentication methods continue to be one of the most common user authentication approaches to online services, but also one of the least secure. These issues have led to seeking safer and more privacy-friendly approaches to managing digital identity. In recent years, one of the most discussed approaches in this regard is Self-Sovereign Identity (SSI). It is a decentralized identity model that allows the user not to delegate the custody and control of personal data to third parties, but the users themselves become the sole owners and managers of their own identity (Soltani et al., 2021).

The introduction of the GDPR in the European Union (EU) countries in 2016 has increased interest in solutions such as the SSI. Indeed, this regulation aims to strengthen the protection of personal data of citizens of the European Union, both inside and outside the borders of the EU, restoring control of their own identity to the citizens. Furthermore, the recent development of some emerging technologies has given a further boost to the growth of the SSI model. Among these in particular there is the blockchain technology. While not necessary for SSI, many believe that such technology can provide the technical foundation upon which the concept of Self-Sovereign Identity can be realized. There are, indeed, multiple mutual benefits in the combination of SSI and Blockchain. This has fueled the excitement where many use-cases for different scenarios are being explored to understand the suitability of such a system (Ferdous et al., 2019).

However, the numerous papers on the SSI topic, as well as the many independent application cases developed, carry an undesirable side effect. That is, the presence of multiple different interpretations of the SSI model, which reflect a certain confusion on the subject. This work aims to shed light on the concept of SSI through an in-depth study of the model. Furthermore, it is proposed to fill the lack of academic articles that delve into the cases of practical application of the model. To this end, a census of SSI cases was carried out as part of this research work. These surveyed projects were then analyzed on the basis of the most relevant variables that emerged following the study on the model, in order to intercept the main trends and common practices in place. To achieve these goals, this work has been structured in:

- Chapter 1: Literature Review contains the analysis of the literature on the subject, providing a theoretical background of the SSI model. All the basic definitions, concepts and tools

related to the model are provided starting from a wide range of articles. At the end of the chapter, the main gaps of the literature are identified, starting from them the research question of this thesis work is defined.

- Chapter 2: Research Methodology explains in detail how this research work was carried out, from the theoretical review to the empirical framework.
- Chapter 3: Results reports the results of the analysis of the cases on the basis of the relevant variables identified following the review of the literature. It also contains the analysis of these results with the discussion of the key messages extracted.
- Chapter 4: Conclusions elaborates the results reported, summarizing the answer to the research question. The chapter also reports the limitations of the work and potential future research directions.

Chapter 1: Literature Review

This chapter contains the results of the literature review work. A systematic review of the state of the art and of the main theoretical concepts underlying this thesis. In particular, the chapter is divided into three parts. The first will discuss the concept of digital identity, showing how it works and its major benefits. Subsequently, the main evolutionary trends related to it will be analyzed. In the second part, blockchain technology and the possible relationships with digital identity will be presented. Then, the concept of Self-Sovereign Identity, the main theme of the thesis, will be dealt with in detail. The chapter ends with the formalization of the literature gaps and the consequent definition of the Research Question. The Research Question is the driving force of this research work. It is directly related to the gaps, identified at the end of the Literature Review process. The Research Question has been divided into sub-questions, which allow to analyze and focus on some specific points of interest. The answers to these questions are the main contribution of the empirical work carried out within this thesis.

1.1 DIGITAL IDENTITY

1.1.1 Definition

Although not all definitions of digital identity are completely in agreement and sometimes focused on distinct aspects, it is possible to note that there are recurring themes and a common basis for all of them. This indicates a good level of maturity on the theoretical concept of digital identity.

Sullivan (2012) stated that "*Digital identity is an individual's identity which is composed of information stored and transmitted in digital form."* And more: "*Digital identity is all the information digitally recorded about an individual, which is accessible under the scheme*." It is a very simple definition, thanks to which it is possible to identify the key point of this concept, namely the digitization of all aspects relating to identity.

The Digital Identity Observatory of the Politecnico di Milano has outlined a definition, aligned with the previous one, but which broadens the vision on digital identity. Indeed, according to the definition "*digital identity is a set of data that permits to uniquely identify a person, company or object, which are collected, stored and shared digitally inside of an ecosystem of actors, through enabling technologies, which allows access to digital services with added value.*" In this case, the definition is broader and focuses on digital identity as a system. Effectively, from this definition it is possible to identify the four pillars of the digital identity system:

1) **Identity dataset**: it includes all the possible data associated with the identity, regardless of their nature which can be static, as in the case of personal data or dynamic, when it comes to the interactions enabled by digital identity.
2) **Ecosystem**: it is made up of the actors participating in the system. Each actor can cover one or more roles depending on the type of model and architecture of the system.
3) **Enabling technologies**: depending on the infrastructural layer in which they are applied, they can be divided into architectural, integration and process technologies.
4) **Added value digital services**: it includes all the value-added digital services that digital identity allows to access, both online and offline.

Also, for the World Bank Group, GSMA and Secure Identity Alliance discussion paper (2016) "*digital identity is a collection of electronically captured and stored identity attributes that uniquely describe a person within a given context and are used for electronic transactions*." While "*a digital identity system refers to the systems and processes that manage the lifecycle of individual digital identities*." In this case, both aspects seen above are dealt with, albeit in a less detailed way. This confirms the existence of consolidated themes and a common basis in dealing with the concept of digital identity.

Enlarging the concept, digital identity is a snapshot of the actual identity of an entity, including other than people, also companies, devices, or cars (Der et al., 2017). Oxford Advanced Learner's Dictionary (2022) defines the entity noun as *something that exists separately from other things and has its own identity*. The World Economic Forum recognizes three different types of entities: individuals, legal entities, and assets. Der et al. (2017) continue their definition by mentioning some limitations in the characteristics of digital identity. They point out that, while possible, digital identity may not necessarily represent all the attributes of a real entity. Furthermore, sometimes it is limited to a purpose (i.e., using a service) or a context (i.e., interacting with other similar entities). In addition, current digital identity has a well-defined temporal context, and needs to be updated, as some information can change over time. For these reasons a single entity, especially humans, could have hundreds of digital identities (Der et al., 2017).

McKinsey Report (2019), instead, gives a more operational perspective on digital identity, stating that: "*Digital ID is verified to a high degree of assurance, unique, and established with individual consent, and protects user privacy and control over data. It authenticates your identity over digital channels through one or more factors.*" Unlike the others, the focus of this definition is on the characteristics of digital identity. It is therefore an almost practical definition, in the sense that it indicates the main attributes that digital identity should possess. These attributes, mentioned in the definition, are detailed below:

- **Is verified and authenticated to a high degree of assurance**: the verification process takes place during the initial registration and involves checking that the information declared defines the identity of the person. Authenticate an identity means demonstrate the

association between an entity and a set of identifiers (Camp, 2004) or an identity previously established. High-assurance digital identity meets both government and private-sector institutions' standards for initial registration and subsequent uses. To achieve unique high-assurance authentication and verification, different possibilities are available, including credentials such as biometrics, passwords, QR codes, and smart devices with identity information embedded in them.

- **Is unique**: an individual has only one digital identity within a system, and every system identity corresponds to only one individual.
- **Is established with individual consent**: individuals register for and use the digital identity with knowledge of what personal data will be collected and how they will be used.
- **Protects user privacy and ensures control over personal data**: built-in safeguards ensure privacy and security, while also giving users access to their personal data, decision rights over who has access to that data, and transparency into who has accessed it. Moreover, according to Beduschi (2019), one of the main potentials of digital identity is to use technology to provide official identification to populations in need. Indeed, digital identity may render individuals without legal documentation more visible and therefore less vulnerable to abuse and exploitation. Several initiatives, both private and public, are active to achieve this goal. All of them have profound implications for the protection of human rights. Digital identity will only effectively contribute to the protection of human rights if it, among other things, promotes high standards of privacy and data protection.

Based on the configuration of the different variables that make up the digital identity, described so far, it is possible to identify five different types of digital identity:

1) **Social ID:** it is composed of the set of data, generally self-declared by the user when registering on a social platform. These data are characterized by a minimum level of verification and a high frequency of updating and enrichment. This type of ID can also be used for access to other digital services with a low level of criticality. The most common examples of this type are the identities created with Facebook or Google (Digital Identity Observatory of the Politecnico di Milano, 2020).
2) **eCommerce ID:** it has similar characteristics to the Social ID model, but is based on eCommerce platforms, such as profiles created on Amazon or Shopify (Digital Identity Observatory of the Politecnico di Milano, 2020).
3) **eGovernment ID:** it includes digital identity systems created and managed, sometimes indirectly, by government agencies. They are used for authenticating and log into e-government services and give a reliable and unique identity to their citizens (Pöhn et al., 2021).
4) **Financial ID:** it is composed of the set of identification data collected by a financial institution, such as a bank, to recognize its customer. They can also be used in single-sign-on

access mode for other digital services. An example is the Swedish BankID system, or even the identity created with the PayPal service (Digital Identity Observatory of the Politecnico di Milano, 2020).

5) **Mobile ID:** this model is based on the use of the SIM card as a secure element for identity data. In this case, the data are generally collected and verified with medium-high guarantee levels. An example is the itsme system (Digital Identity Observatory of the Politecnico di Milano, 2020).

1.1.2 Functioning and main differences with the traditional model

To introduce the functioning of digital identity, its lifecycle is described using an example created by the World Bank Group (2018) and aligned to what described in the USAID report (2017), is presented in *Figure 1.1*.

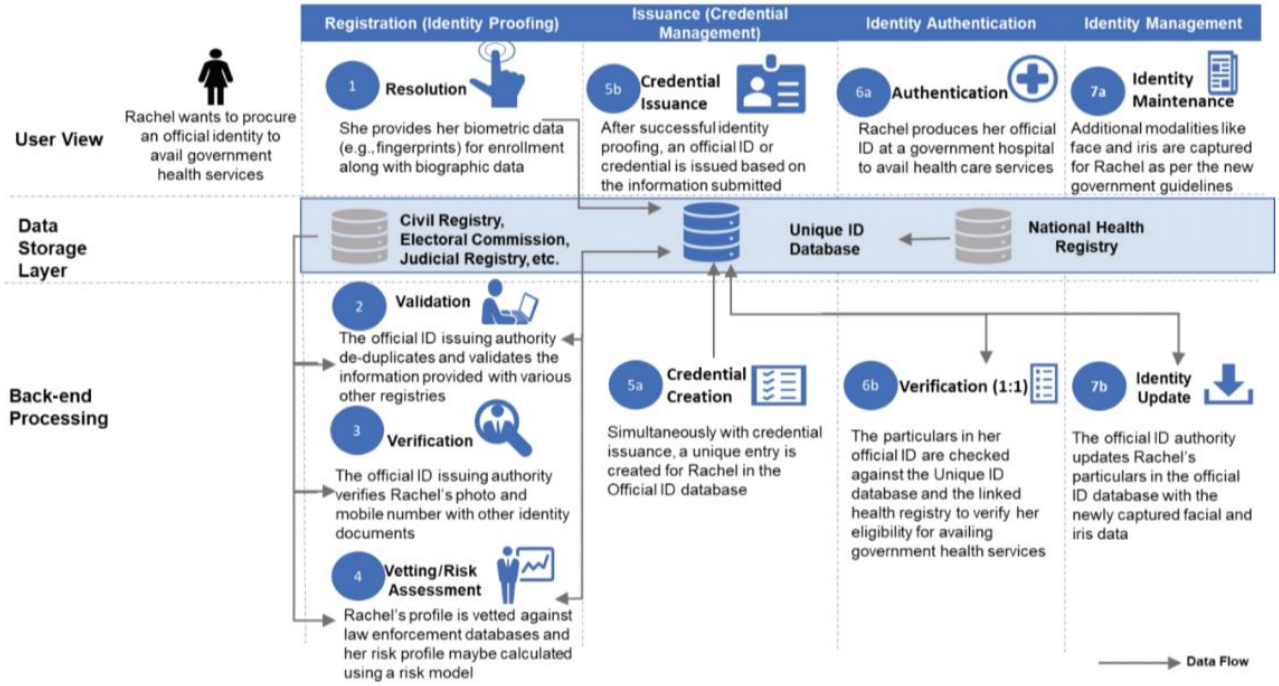

Figure 1. 1 Rachel's Journey through the Identity Lifecycle (World Bank Group Report, 2018)

The lifecycle of an identity, as can be seen in the example, starts when a person applies for a digital identity and ends when the ID is invalidated due to death, request for removal by the individual, or some other event (World Bank Group Report, 2018).

The first phase is **registration** or **identification**, which is the most crucial step in creating a digital identity (World Bank Group et al., Discussion Paper, 2016). Generally, registration happens where a digital identity is required to enable an experience. Key identity attributes are captured from a user

as part of his journey (Cameron & Grewe, 2022). This phase is in turn divided into other sub-phases. The process starts with **resolution** or **enrollment,** the user provides the issuing authority a series of information (i.e., biographic information, breeder documents, photographs, etc.). The information presented varies depending on the type of digital identity and its required Level of Assurance, a concept elaborated on later in this section. Based on that, information can be provided in person or online. In-person proofing is required for the highest identity assurance level, IAL3 (National Institute of Standards and Technology, U.S. Department of Commerce, 2017). So that the information can be validated and augmented by the registration authority as needed (World Bank Group Report, 2018). The next steps are **validation**, where the authority determines the authenticity, validity, and accuracy of the identity information provided, and relates it to a living person. Followed by **verification**, here a link between a claimed identity and the real-life subject presenting the evidence is established. The last step is **vetting** or **risk assessment**, where the user's profile is assessed against a watch list or a risk-based model (World Bank Group Report, 2018). Registration interactions are typically one-time with the customer. Generally, the registration phase concludes with a confirmation of the purpose of the flow (Cameron & Grewe, 2022). This first phase is also called onboarding.

The second phase is **credential management**. This part starts with **credential issuance**, which is the process of creating and distributing virtual or physical credentials (World Bank Group Report, 2018). Here it is possible to observe the first difference with the traditional process, indeed, conventionally identity issuers provide printed documents or credentials (i.e., birth certificate, identity documents and passports, etc.) (Atick, 2016). Instead, in digital identity systems, the credentials or the certificates issued must be electronic, in the sense that they store and communicate data electronically (i.e., smartcards, 2D barcode card, mobile identity and ID in the cloud) (World Bank Group et al., Discussion Paper, 2016). Credential management includes also other two parts: **maintenance,** which is the retrieval, update, and deletion of credentials. And **revocation,** namely the removal of the privileges assigned to credentials (World Bank Group Report, 2018).

Once a person has been registered and credentialed, it can use its digital identity to access the associated benefits and services (World Bank Group et al., Discussion Paper, 2016). For example, in *Figure 1.1* Rachel wants to use her identity document to have access to healthcare service in a hospital. Before accessing any services, she, like all the other users, must validate her credential (Cameron & Grewe, 2022). This phase is called **identity authentication**. It is the process of verifying an identity claim against the registered identity information (World Bank Group Report, 2018). Such information is called Authentication Factor and can be divided into four categories:

- **What a person is**: intrinsic user information, such as biometrics or biographical data.
- **What a person knows**: information that is theoretically in possession only of the person concerned. It is the traditional method of authentication, it includes, indeed, password and PIN.

- **What a person has**: in this case the authentication takes place thanks to something that the user has. It can be either tangible, like a document or badge, or intangible, like a security token.
- **What a person does**: the specific behavior of the person allows its recognition by particular advanced systems.

The most common Authentication Factors are summarized in *Figure 1.2* (World Bank Group et al., Discussion Paper, 2016).

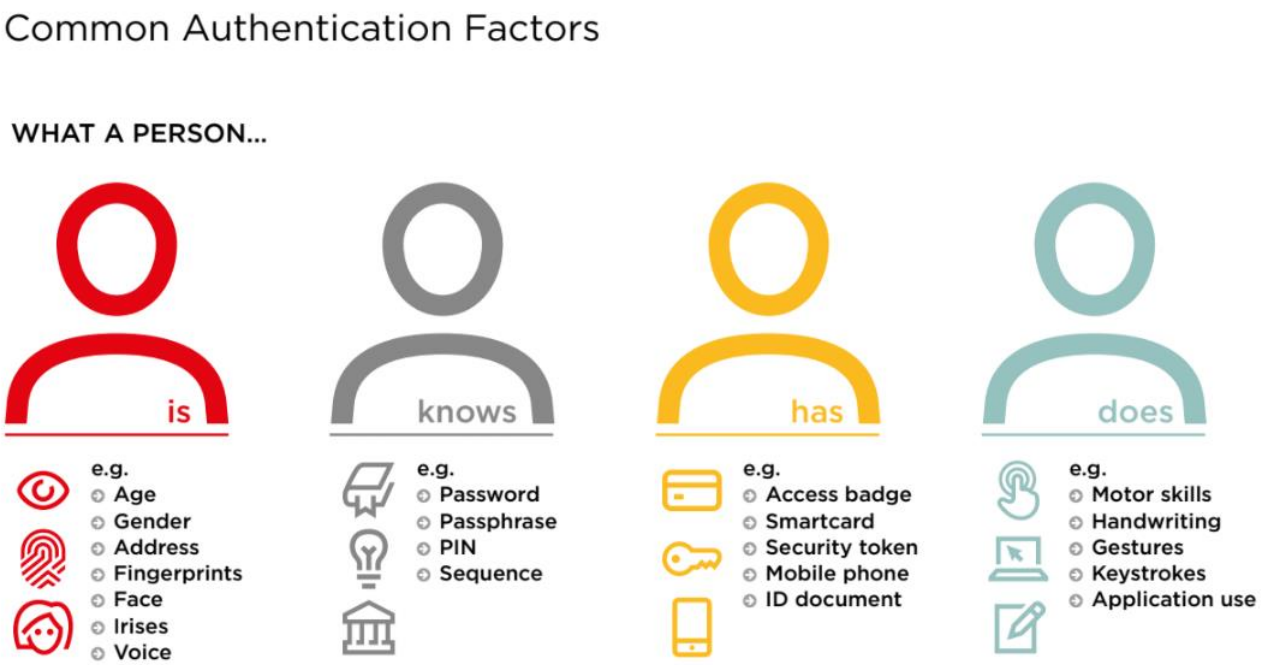

Figure 1. 2 Common Authentication Factors (World Bank Group et al., Discussion Paper, 2016)

If two or more independent authenticators, from at least two distinct categories, are required, the process is characterized by Multi-Factor Authentication (MFA). Otherwise, it is a Single-Factor Authentication (Bertino et al., 2007). Using MFA provides more security and is often critical for systems that require stronger authentication (Ometov et al., 2018).

The following phase is **authorization,** which takes place after the authentication of an individual's claim of identity. It defines access rights that a relying party or Service Provider (SP) has associated with the identity, independently of the Identity Provider (IdP) (i.e., the National Identification Authority). There are also more advanced authorization schemes, where access rights are contextual and dynamic (World Bank Group Report, 2018).

Lastly, another important part of the digital identity lifecycle is **identity management**, which similarly to credential management, is composed of **maintenance** and **revocation**. More specifically, identity management consists of retrieving, updating, and deleting the attributes of the identity or data fields and policies, which regulate user's access to information and services (World Bank Group

Report, 2018). Identity management allows users to keep the status and content of their digital identity up to date. Updating some attributes is faster and more efficient if individuals use a digital identity system. With the traditional method, if a person wants to change the address, for example, it would be necessary to replace the entire physical document with a new one. Instead, thanks to digital identity, it is sufficient to update just the relevant field (World Bank Group et al., Discussion Paper, 2016). Furthermore, databases must be updated to reflect major life events, such as birth and death (World Bank Group Report, 2018).

As already mentioned above, when a person identifies or authenticates herself using one or more identity attributes, the importance of the validity of the attributes provided changes depending on the degree of security needed and the context in which the information is acquired. This is called **Level of Assurance** (LOA). Assurance levels depend on the strength of the identification and authentication processes. They are critical to access control and decrease the risk of identity theft. A thorough discussion of the subject would require a lot of space and is beyond the scope of this work. However, *Figure 1.3* shows the key elements and characteristics of the various levels. As a general guideline, it is possible to say that the higher the LOA, the lower is the risk that SP will rely on a compromised credential during a transaction. Moreover, distinct types of services will require different LOAs, not always the highest level will be required. However, an elevated level is usually required to access essential and highly personal services and or information (World Bank Group et al., Discussion Paper, 2016).

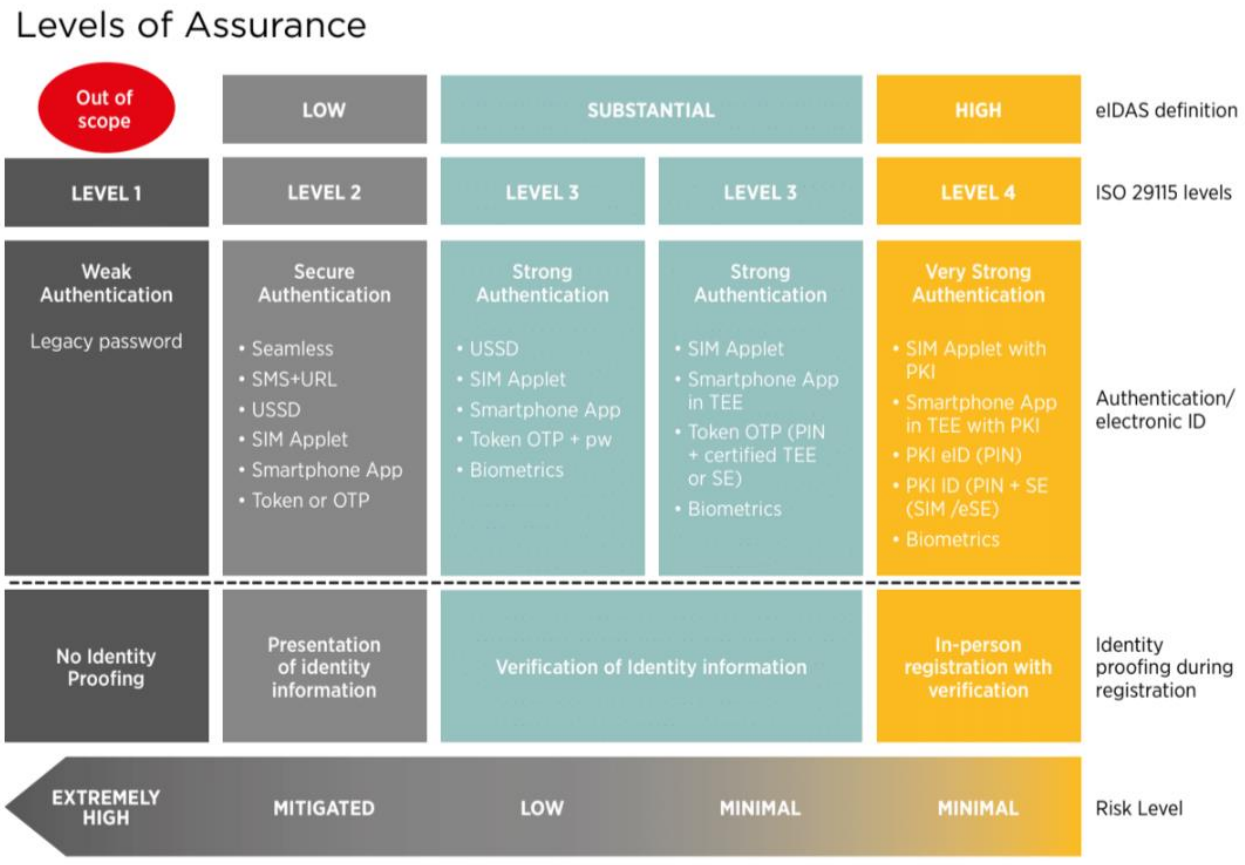

Figure 1. 3 Level of Assurance recap (World Bank Group et al., Discussion Paper, 2016). Key: OTP = one-time password; PKI = public key infrastructure; (e)SE = secure element or embedded secure element (a tamper-resistant hardware platform); TEE = trusted execution environment (a secure area of the smartphone); USSD = unstructured supplementary service data ("quick codes").

Within the digital identity ecosystem, it is possible to identify different actors, covering specific and various roles. There is not a defined and common set of roles, as they also depend on the type and purpose of the digital identity. However, according to World Bank Group et al. (2016), it is possible to recognize some recurring roles within the system and divide them in three large categories:

1) **End-Users**: individuals are the end-users of digital identity systems. They create a digital and use the credentials they receive to access the services of a given country or company.
2) **Providers**:
   - **Government bodies**: institutions develop policies and legal frameworks to enable the acceptance of digital identity. At the same time, they create regulations to protect user privacy and all other fundamental rights. Governments can collaborate between themselves and with international bodies to develop a universal standard, and also with the private sector to understand the economy of digital identity and how to enhance it within the public-private model (McKinsey Report 2019). Within this group we find three sub-categories of actors. The first includes Legal Registrars, which are the agencies in charge of providing legal identification to citizens, such as the national identification authorities. The second includes Functional Registrars, which are agencies that create and maintain identity registries for a specific purpose or service. Some examples are electoral commissions, tax agencies, social security authorities and hospitals. These registries can be linked to Legal Registries, but they can also be separate identity systems. Lastly, there are eGovernment SPs, which are government agencies that provide online services to citizens, which require some proof of identity. They are often linked to the two previous sub-categories.
   - **Private firms**: private organizations can innovate processes that could leverage digital identity to boost efficiency and improve customer experience. Like governments, they can work to facilitate development of global standards, and to conduct analysis of digital identity, developing new digital identity programs (McKinsey Report 2019). In this case it is possible to distinguish two sub-categories. The first includes Commercial Service Providers, which are companies that either use digital identities to provide services to their clients or enable end-users to transact in a digital environment providing digital identity and authentication services. In the second sub-category, on the other hand, there are the Identity Solution Suppliers, which are firms that provide hardware, software, and technical support for the development of digital identity systems.

- **Digital Identity Providers**: these actors create digital identities for users by registering them and issuing documentation or credentials. In general, IdPs also store and manage data and credentials on behalf of the users. IdPs can be both in the public and private sectors. Even if oftentimes, private IdPs rely upon or use legal identity provided by the public sector.
- **Attribute Providers**: these entities hold verified user data and either verify or provide these attributes to third parties. Such information pertains to various areas of the individual's identity data. In many cases, there is an overlap between IdP and Attribute Providers. However, in some cases actors provide attributes upon request of the IdPs or other relying parties.
- **Digital Authentication Providers**: they verify a user's attributes or identity in order to determine his or her right to access a service or benefit. In the public sector, those agencies that are directly involved in delivering services that require verification are commonly also Authentication Providers. In the private sector, generally, Commercial Service Providers authenticate users.
- **Service Providers**: these entities provide services directly to end-users. They can be public agencies, such as functional registrars and eGovernment SPs, as well as private SPs. SPs may themselves be IdP and Authentication Providers, or they may outsource these functions to other agencies.

3) **Enabling and Supporting actors**:

- **Regulatory and oversight agencies**: organizations that regulate, control and audit digital identity systems. The goal of these actors is to ensure that IdPs and authentication providers follow legal standards and best practices for the collection, storage, and use of personal data. Basically, they assure a consistent identity management, also supervising and legislating on issues such as data protection, privacy, security, and user trust. An example of this type of agency is the European Data Protection Board.
- **Standard setting bodies**: organizations that provide protocols for digital identification and authentication. The goal of these agencies is to increase interoperability and build open, robust, and scalable identity solutions. Some examples are NIST and the Open ID Foundation.
- **Identity organizations and trust frameworks**: entities that define technical, operational, legal, and enforcement mechanisms for information exchange related to identity management. Their goal is to establish trust among the stakeholders in the system. Some examples in the public and private sectors, respectively, are Trust Framework Provider Adoption Process and Mobile Connect.
- **Donor agencies and development partners**: entities that provide support to activities related to digital identity. Either through funding or providing technical

support. Their goals, as well as their motivations, are often varied, but they generally try to support projects deemed most valid or those implemented by the government to achieve specific goals, which often coincide with global development goals.

In addition to the peculiarities of digital identity discussed so far, the main differences between digital and physical identity, according to the Digital Identity Observatory of the Politecnico di Milano (2020), will be reported below:

- **Proliferation**: physical identity is generally unique or linked to a few valid and recognized identity tools, associated with a physical document (i.e., identity card, driving license or passport). Instead, each individual can have a lot of digital identities, as there are many online systems and platforms that, with various levels of reliability, allow you to create a profile of interoperable data within an ecosystem. So, it is possible to activate and own multiple digital identities at the same time. Users face the problem of managing all these various digital identities, including those of social networks, those linked to national systems and those in the business sector.
- **Validity**: physical identification documents are typically accepted throughout the national territory and in many cases are also valid at an international level. Digital identity, on the other hand, is recognized only within the ecosystem of actors who have decided to join the system and to adopt the appropriate technological infrastructure to integrate with the digital IdP.
- **Ecosystem**: in the case of physical identity, the entity that issues the identification document is not involved in the subsequent interactions in which it will be exhibited. Instead, for digital identity, at each interaction the actors involved initiate the exchange of the identification data necessary for the provision of the service requested by the user. In the base case these actors are user, SP and IdP. To access an online service or a third-party app using the Facebook Login, for example, it will be necessary to pass data between the social network and the SP. So IdP is continuously involved in the case of digital identity.
- **Dynamicity**: physical identity has a predefined and static set of individual identification data, including personal data and sometimes biometric data. The data that, on the other hand, constitute the digital identity, are more dynamic and generally updated with higher frequency. Furthermore, digital identity can be enriched with different and specific information depending on the context (i.e., legal, health, financial, etc.). The dynamism and richness of the data that make up digital identity represent the greatest potential opportunities for value creation.

Lastly, to better understand the benefits of digital identity, it is first necessary to make a brief excursus on the current situation. Indeed, as of 2018, the ID4D Global Dataset estimates that there

are one billion people worldwide who do not have basic identity documents[1]. The lack of legal identity results in denial of essential rights such as voting, having access to banking, education, and health services. With the consequent difficulty in finding work and legal recognition (Nyst et al., 2016). Also, according to the ID4D Global Dataset[1], the problem mainly concerns the poorest countries in sub-Saharan Africa and in southern Asia. And in turn, it affects the poorest and most vulnerable groups of people. From this sample, it emerges that the lack of a recognized identity afflicts about half of the female population in low-income countries, thus limiting their access to critical services and participation in political and economic life. Furthermore, Multiple academic studies (Clark & Gelb, 2013; Gelb & Metz, 2018; Muralidharan et al., 2020) demonstrate the importance of having a robust identity to reach various development goals. For these reasons, in 2015, the United Nations (UN) General Assembly recognized the importance of having a legal identity, defining a specific Sustainable Development Goal (target 16.9: "free and universal legal identity, including birth registration, by 2030.").

The situation is not perfect for the remaining people with an identification either, as more than half of them cannot effectively use it in today's digital world. This prevents them from having access to the digital economy and the full achievement of their social and humanitarian rights (McKinsey Report 2019).

What has been said so far explains why an adequate digital identity system can unlock a significant amount of value, estimated around 3% GDP equivalent per-country for developed economies in 2030 and 6% GDP equivalent for emerging economies. This value can be divided equally between individuals and companies / government institutions (McKinsey Report 2019).

Going into detail, it is therefore possible to distinguish and analyze the main benefits for the two types of entities. The main benefits for individuals are:

- **Access to essential social assistance services and human rights recognition**: This is a problem that affects especially developing countries. The lack of valid identification documents causes difficulties (or makes it impossible) to access essential services, as well as the recognition of individuals' humanitarian rights (Nyst et al., 2016). For example, in Kenya the lack or the misrecognition of identity documents of migrants and refugees prevents them from accessing basic social assistance services (Weitzberg, 2020). In recent years there has been an exponential increase in access to the internet[2] and digital technologies in developing and least developed countries. This, according to the McKinsey Report (2019) and The World

[1] https://id4d.worldbank.org/guide/why-id-matters-development

[2] https://www.statista.com/statistics/209096/share-of-internet-users-in-the-total-world-population-since-2006/

Bank[3], could favor the spread of digital identities, helping in resolving the problem mentioned above.

- **Increased use of financial services**: digital identity helps individuals meet Know Your Customer (KYC) requirements and enables remote customer registration for financial services. Removing the problems of lack of documentation, distance to financial institutions, and cost of financial services, favors the access to financial services and credit lines (McKinsey Report, 2019).
- **Improved access to employment**: digital identity enables better digital talent matching and contracting platforms, which allow job seekers to authenticate themselves online. Such platforms could facilitate access to labor markets for inactive and unemployed workers. Moreover, it could also boost labor productivity (McKinsey Report, 2019).
- **Greater agricultural productivity from formalized landownership**: digital identity could help improve incentives to make larger and longer-term investments in farming, thanks to the possibility of formal land titling. Digital identity could also bring benefits to farmers through better targeting of agricultural support, especially when combined with location information and remote sensing (McKinsey Report, 2019).
- **Time savings**: digital identity enables the digitization of sensitive identity-related interactions, which allows the simplification and automation of identity related processes. At the same time, it reduces the need for travel, a particular benefit for people who live in rural areas. A similar effect is also foreseeable with regard to the completion of tax-related operations, also saving time for tax departments in processing and auditing (McKinsey Report, 2019).

The last four points are the largest contributors to direct economic value creation for individuals, enabled by digital identity (McKinsey Report, 2019).

As far private and public institutions benefit goes, the main ones are:

- **Time and cost savings**: institutions using high-assurance digital identities could see up to 90 percent cost reduction in customer onboarding, with the time taken for these interactions reduced from days or weeks to minutes (McKinsey Report, 2019).
- **Reduced fraud**: digital identity can help reduce fraud in a wide range of transactions. From decreasing payroll fraud to reducing identity fraud (McKinsey Report, 2019).
- **Increased sales of goods and services**: digital identity could improve customer experience in digital channels. This would lead to an increase in sales and gains for companies, especially for those who rely on high-assurance identities, such as banks and digital gig economy platforms (McKinsey Report, 2019).

[3] https://www.worldbank.org/en/news/immersive-story/2019/08/14/inclusive-and-trusted-digital-id-can-unlock-opportunities-for-the-worlds-most-vulnerable

- **Greater employment and labor productivity**: as mentioned in the part of individual benefits, digital identity offers a series of gains for jobseekers. Therefore, there is also an impact on institutions, indeed, businesses could more rapidly fill open positions and find the right employee for a given position, leading to higher productivity (McKinsey Report, 2019).
- **Better tax collection**: greater revenue facilitated by the use of digital identity, could expand the tax base, helping promote formalization of the economy and more effective tax collection. Emerging economies, in particular, could experience substantial benefits, if they first make it an explicit goal and then build the requisite tax collection tools enabled by digital identity programs (McKinsey Report, 2019).

Similarly, to the previous case, these five points are the largest sources of value for institutions, in both government and the private sector, enabled by digital identity (McKinsey Report, 2019).

However, to be able to fully create value, there are some key points on which pay particular attention. Among these, the main ones identified are:

- **Inclusion**: first, ensure universal coverage for individuals from birth to death, free from discrimination. And second, remove barriers to access and usage, such as direct and indirect costs, disparities in the availability of information and technology (World Bank Group Document, 2017; World Bank Group Report, 2018).
- **Design**: establishing a unique, secure, and accurate identity. Creating a platform that is interoperable (World Bank Group Document, 2017; World Bank Group Report, 2018). The importance of interoperability in digital identity systems is shown in different academic research (Gasser & Palfrey, 2007; Rundle & Trevithick, 2007; Gelb & Metz, 2018), which also highlight several advantages for systems incorporating interoperability. The designers should also consider using open standards and preventing vendor and technology lock-in effects. Furthermore, Identification systems should be designed for long-term economic and operational sustainability. Lastly, system design can (and has to) assure control and protection of user privacy, which is the next point on this list (World Bank Group Document, 2017; World Bank Group Report, 2018).
- **Privacy**: the rising concern of people around privacy, as shown for example by KPMG[4] and Internet Society[5] reports, makes the safeguard of data privacy, security, and user rights fundamental. To succeed in this, a legitimate and comprehensive legal and regulatory framework should be realized. Consequently, clearness in governance aspects, such as institutional mandates and accountability, must be established. Lastly, to ensure the correct functioning of the system and quickly resolve any problems or disputes, the legal and trust

[4] Available at: https://advisory.kpmg.us/articles/2021/bridging-the-trust-chasm.html

[5] Available at: https://www.internetsociety.org/wp-content/uploads/2019/05/CI_IS_Joint_Report-EN.pdf

frameworks must be enforced through independent oversight and adjudication of grievances (World Bank Group Document, 2017; World Bank Group Report, 2018).

- **Integration**: which can be divided into two categories. Integration with emerging technologies and with other compatible services. In the first case, the use of new technologies can improve the current system or even facilitate the creation of a new one. This is the case, for example, of the blockchain, which, as mentioned in various academic papers, can solve some of the problems of digital identity systems, and also favor the adoption of the new Self-Sovereign Identity model (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020; Gstrein et al., 2020). In the second case, digital identity creates opportunities to improve other existing services and vice versa (Atick, 2016).

1.1.3 Evolution of digital identity

The landscape of identity management has gone through an evolutionary path (*Figure 1.4*): starting from the traditional model, the centralized one, and then evolving in distinct phases with the introduction of new models. The four models can coexist, but are often thought of as a progression.

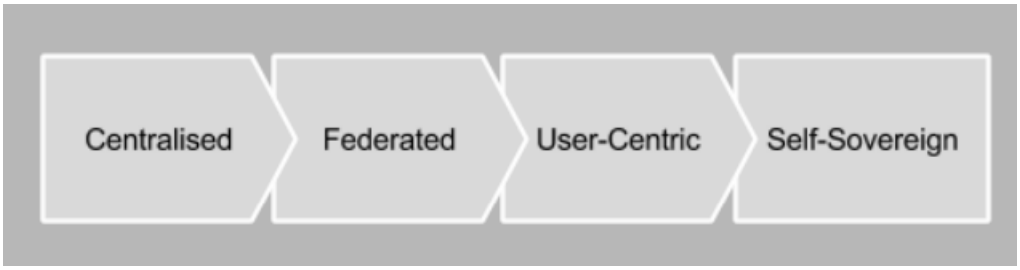

Figure 1. 4 The evolution of digital identity (Sovrin White Paper, 2016)

1.1.3.1 Phase One: Centralized Identity

The Centralized Identity model is the more traditional and most widespread paradigm. In this case, an individual accesses the services of an organization that manages or owns the identity system. The owner of the system collects, stores, and uses the individual's identity and its related data. Such systems are currently proposed by various private organizations such as banks, social media companies, and even governments (Dib & Toumi, 2020). In Centralized systems, every user interaction with SPs must be authenticated through the central IdP (Jøsang & Pope, 2005; World Economic Forum, 2016). More specifically, the users enter their data and, if necessary, the related proof in the system. The central entity verifies the data and, in the event of a positive outcome, issues a valid and certified digital identity. Subsequently, to access the services offered by a SP, users must authorize the latter to access their data, held by the central IdP. The SP, with users' consent,

will thus be able to access and verify the user's identity in the central database and, if the requirements are met, it will grant the applicant the possibility to access the services (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020).

Despite being the most widespread model and the one which guarantees greater control to the entity issuing the identity, the Centralized Identity model has some major disadvantages, which often affect the end user. Indeed, the Centralized model is susceptible to various security attacks, with potentially disastrous consequences (Soltani et al., 2021). Indeed, a breach in the central entity or its malfunction causes a fall of the entire identity system, this phenomenon is called Single Point of Failure (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020). Furthermore, since data are managed by a centralized third party, the privacy of users may be compromised, and their online activities may be linked and eventually traced (Dib & Toumi, 2020). In this model the individual does not have full control over the use of their data. Individuals are locked into a single authority who can, although highly unlikely, even deny or falsify their identity. Centralization naturally empowers the central entities, not the users. So, they have no choice but to trust the identity manager. In addition, there is also the risk that the central entity knows all the data of all the participants, the so-called Big Brother Effect. (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020). Another problem with this model is that the consumers will have to create one identity per SP, with the result of ending up with numerous partial identities, which become increasingly difficult to manage (Dib & Toumi, 2020; Ferdous et al., 2019). From the point of view of the IdPs, centralized models require investing high resources to store, maintain, and protect users' data (Pöhn & Hommel, 2020).

An example of the centralized identity model is the Aadhaar system, set up by the Indian government. It is a unique identification number, which is assigned to each citizen who requests it. In order to obtain it, each applicant must provide a series of biographical and biometric data, which are recorded and stored in a central system. Aadhaar was designed to avoid the proliferation of fake identities and simplify access to digital services that require authentication. This data is centralized and there is no need to re-check them for each SP (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020). Aadhaar suffered numerous attacks, which resulted in the theft of several identities of Indian citizens (Dixon, 2017; Ishmaev, 2020). The largest of these beaches was in 2018 and affected more than a million users according to the World Economic Forum's Global Risks Report (2019). This is an example of Single Point of Failure.

Despite all these drawbacks, as previously said, identity on the Internet today is still centralized. However, in the last years there has been an attempt to return identities to the people, so that they could control them.

1.1.3.2 Phase Two: Federated Identity

In the federated identity model, a set of SPs and IdPs forms a trusted federation. This allows the individual to have the possibility to select a single IdP, among those federated. Once the IdP has been selected, it can be used to access the services of any of the participating SPs (Soltani et al., 2021). Users often like the convenience of this model, which has led to a widespread adoption of Federated systems, especially when the identity data is only shared between trusted entities, like in governmental services and international organizations (Ferdous et al., 2019). The federated identity model allows a user to use a single set of credentials to authenticate with the IdP and then access any of the SPs present in the ecosystem. This is achieved through single sign-on, where the IdP provides an authentication token to the SP (Chadwick, 2009). The first difference with the Centralized model is that the data is stored on several different databases, no longer on a single centralized one. This solves the problem of the Single Point of Failure. However, the other issues for users remain similar to those seen in the case of Centralized Identity. Moreover, also on the other side there are problems, indeed, building trust relationships between two or multiple system owners is complicated and limits the fast implementation of the system. In particular, the complexity for system owners arises from the eventual need for legal agreements, division of costs and risks, and the creation of technical standards. All this leads to high implementation costs (Dib & Toumi, 2020).

An example of Federated Identity model is any system based on Security Assertion Markup Language (SAML) protocol (Soltani et al., 2021). SAML standard defines a framework for exchanging security information between online entities. It expresses assertions about a subject in a portable fashion that other applications in the system can trust (Hughes & Maler, 2005).

1.1.3.3 Phase Three: User-Centric Identity

In the User-Centric Identity model users have some liberty in selecting the IdP and the identity attributes they want to share, along with the conditions under which those attributes should be shared (Jøsang & Pope, 2005). Thanks to the significant role played by interoperability and consensus in these systems, it is possible to overcome most of the problems of the two previous models. However, true control of the data is not yet completely in the hands of the user (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020).

OpenID Connect (OIDC), based on Open Authentication (OAuth) 2.0 protocol, is the dominant technological protocol[6] in this model. OIDC permits to verify the identity of the end-users based on the authentication performed by an Authorization Server, as well as to obtain basic profile information about them. An example is Google or Facebook login. In this model, the SPs and the

[6] An authentication protocol is a type of communications protocol specially designed for transfer of authentication data between two entities. It specifies the type of information needed for authentication as well as rules, syntax, semantics and possible error recovery methods.

IdPs may not always have built a direct trust relationship. SPs rely on the fact that the other party has verified the user's identity. Basically, it is possible to access to their service logging in with a third-party account, with which a direct relationship of trust has not been created but is in any case considered dependable. This authentication method is widely adopted. However, in this way, IdPs can trace the users' activities, obtaining confidential information (Sakimura et al., 2014; Ferdous et al., 2019; Dib & Toumi, 2020).

It is therefore clear that to bring identity back to people being user-centric is not enough.

1.1.3.4 Phase Four: Self-Sovereign Identity

The Self-Sovereign Identity (SSI) model will be explored in detail in Chapter 1.3. Here the focus will be on framing this scheme within the evolutionary framework of digital identity.

The User-Centric Identity model has brought attention to interoperability and user consent, while maintaining centralized control. It is a major step toward user control of identity, but just a step. Self-Sovereign Identity is the next step to ensure user autonomy, not available in centralized and federated identity models. In the SSI model, the identity holders have full control over their data and decide how their data should be shared with others. The SSI model preserves the right for the selective disclosure of the user's data in different contexts (Soltani et al., 2021).

Similar ideas were already being expressed in the early 10's. Loffreto (2012) introduced the concept of "Sovereign Source Authority" by stating that people have a natural right to identity, which was taken away by their own nations. Instead, the birth of the term Self-Sovereign Identity dates to 2016. It was used for the first time in a post by Christopher Allen, an expert in digital identity. The SSI model, as expressed by Allen, was not conceived as a technical solution, but as a theoretical concept of digital identity where the users are placed at the center of the model in full control of their identity (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020). The SSI model represents, at least theoretically, a solution to all the shortcomings of existing identity management models listed until now (Soltani et al., 2021).

Examples of SSI protocol are Sovrin and Jolocom, both of which rely on blockchain technology and Software Development Kit (SDK)[7]. Another widely used integration protocol, in this case, is the Application Programming Interface (API)[8].

[7] SDK is a set of tools that allows the development of software or firmware for a specific platform. Some SDKs are available for free and can be downloaded directly from the protocol creator site, allowing ecosystem partners to use the solution proposed (Dib & Toumi, 2020).

[8] It is a set of commands and objects that allow developers to interact more easily with a program or service, facilitating the interoperability among different systems (Digital Identity Observatory of the Politecnico di Milano, 2020).

1.2 BLOCKCHAIN TECHNOLOGY

1.2.1 Introduction

A blockchain-like system was first theorized in 1982 by David Chaum. More research and publications followed over the years. However, the first blockchain was conceptualized by Satoshi Nakamoto in 2008. He implemented the technology the following year as a core component of the cryptocurrency Bitcoin, where it serves as the public ledger for all transactions on the network. Interest in the blockchain has grown considerably in recent years and several platforms have established themselves, among these the best known are Ethereum, R3 and the Hyperledger project (Sarmah, 2018; Sherman et al., 2019).

According to Bernal Bernabe et al. (2019) "*the blockchain is a public ledger distributed over a network that records transactions (messages sent from one network node to another) executed among network participants. Each transaction is verified by network nodes according to a majority consensus mechanism before being added to the blockchain. Recorded information cannot be changed or erased and the history of each transaction can be re-created at any time*".

From the definition emerge some key characteristics of the technology. From the first part of the definition, **disintermediation** and **decentralization** properties emerge. Indeed, the blockchain allows transactions to be carried out without central third-party intermediaries. Furthermore, the information is recorded on a distributed ledger, i.e., among several nodes, guaranteeing the safety and resistance of the system. For example, Bitcoin allows the exchange of value between entities, which may not even know each other, without going through a trusted central entity (e.g., banks, government institutions, etc.) (Segendorf, 2014). In particular, in the absence of trusted centralized authority, cryptography and consensus mechanisms, which are explained in detail in Chapter 1.2.2, are used to build the necessary trust for the functioning of the system. A widely used consensus mechanism is Proof-of-Work: when a user initiates a transaction, participants try to solve a complex problem to verify it, the first to resolve the problem get a reward and validate the transaction (Liu et al., 2021).

Another important blockchain's property emerging from its definition is **immutability**. Indeed, blockchain, as the name implies, is made up of blocks containing transactions, which record the changing states of data, verified by the blockchain network. Each of these blocks is cryptographically linked to the previous one in chronological order. This linking process makes the chain immutable. As such, if the data is tampered with, the blockchain will crash and the changed point would be easily traced. This is a peculiar feature of the technology. Indeed, in traditional databases, information can be easily changed or deleted. Moreover, the decentralized nature makes it highly

difficult to alter transaction history. Indeed, blockchain transactions are stored in a fully decentralized peer-to-peer network, which replicates data storage, minimizing the risk of data loss or modification. Put simply, if someone wanted to change a certain block, in addition to forcing that, they would also have to change all the proceeding blocks of the blockchain. In addition, because of decentralization it would have to do it on all the ledgers of the network, which could be millions at the same time (Houtan et al., 2020).

Lastly, from the last sentence of the definition above, further characteristics of the blockchain are deduced, namely **transparency** and **traceability**. However, these properties depend on the governance of the platform. Indeed, they can be restricted to a group of actors or even absent. If a blockchain is open source and the functioning of its algorithms is clear and well known to all participants, the transparency property is satisfied. Similarly, if the transfers made in the blockchain network can be traced by all the users, the traceability property is respected. This latter property is often used in the industrial field to certify and track products along the different phases of the supply chain (Cocco et al., 2021).

Blockchain systems can be divided into two categories: **public**, accessible to anyone who wants to connect to the network and view the Blockchain and **private**, which is accessible only to the authorized entities. Furthermore, there are two types of blockchain: **permissioned** and **permissionless**. A permissioned blockchain provides writing permission only to a subset of entities, decided by the network holder. Instead, a permissionless blockchain allows for anyone to write to the blockchain (Soltani et al., 2021).

Public and permissionless blockchains have the advantage of being completely decentralized and independent of any organization. Indeed, all nodes can take part in the consensus process. Instead, in the private and permissioned case, only one or a certain set of nodes takes part in the validation process. In the private case, therefore, at least on a theoretical level, consensus passes through the organization or set of nodes in command. The underlying risk is to return to some sort of hierarchical or even centralized structure. On the other hand, the smaller number of validators allows for a faster and more efficient process, reducing the time to propagate transactions and blocks. Considering immutability, it is nearly impossible to tamper transactions in a public blockchain. Differently, transactions in a private blockchain could be tampered with more easily as there is only a limited number of participants (Zheng et al., 2017). Additionally, permissioned platforms are often developed in accordance with current legislation, while in the permissionless blockchain case it is not always clear how they fit into regulations (Bernal Bernabe et al., 2019).

Lastly, there is a mixed case, namely public and permissioned blockchain platforms. In this case, the system is open for all to read or change the state of the ledger, but the network of nodes performing consensus is permissioned. This kind of blockchain allows for only a restricted and chosen group of participants to write in the ledger.

As mentioned, the best-known application of blockchain technology are cryptocurrencies, however, there are other applications, such as:

- **Timestamp**: multiple academic research (Gipp et al., 2017; Estevam et al., 2021) propose to use Blockchain to certify the date of a document and the guarantee that it has not been modified over time. All the most famous blockchain platforms have their own timestamping system for example to store the date and time when the block is mined. This includes Bitcoin and Ethereum.
- **Token**: digital assets that can be exchanged on a blockchain (Bernal Bernabe et al., 2019). They can represent digital or physical goods or even a right, such as vote, ownership or access to a service (Li et al., 2019). Also in this case, the major blockchain platforms allow the exchange of tokens of various kinds. Starting from Stablecoin, digital assets whose value is pegged to a reference asset. Therefore, their price is stable. Coming to Non-Fungible Tokens, not divisible tokens with a unique identifier.
- **Smart Contract**: set of instructions known to all, which are automatically executed by a blockchain, after the occurrence of a trigger event or in general upon the occurrence of certain conditions, also known previously (Houtan et al., 2020).
- **DApp**: decentralized applications are special applications on a blockchain platform, visible and accessible to all participating nodes. They are created with a standardized language that makes them interoperable and composable. So, they can be combined to create other DApps, giving life to a real ecosystem (Houtan et al., 2020). Popular DApps include Splinterlands based on the Ethereum blockchain and PancakeSwap built on top of the Binance Smart Chain.

1.2.2 Functioning and potential benefits

Blockchain works via a multistep process, which starts when one participant requests a transaction. Transactions are signed by the applicants and submitted to the peer-to-peer network, where they are transmitted to each node. Individual nodes receive the request and validate the transaction, often in a process called Proof-of-Work (PoW), for which they are rewarded. Once validated the transactions are added to a block and stored on the distributed ledger (Di Pierro, 2017).

In a block, in addition to a certain number of transactions, there are also a timestamp, the hash value of the previous block, and a nonce, which is a random number for verifying the hash (Nofer et al., 2017).

A hash is the result of the application of a hash function to a string. The hash function transforms a string of arbitrary length into a fixed length output. The fundamental characteristic of these

functions is that they are not invertible. It is impossible to trace the original string starting from the hash. Furthermore, the hash functions are consistent, so that each hash is associated with a string. At the same time, the probability that two different strings have the same hash is almost nil, effectively guaranteeing uniqueness (Di Pierro, 2017). In this way it is possible to guarantee the integrity of the entire blockchain, since each block is not only verified, but is also uniquely linked to the previous ones up to the first block of the chain, called Genesis Block (Bernal Bernabe et al., 2019). The uniqueness of hash values can prevent fraud since changes in a single block of the chain would immediately change the respective hash value, resulting in a hash mismatch of the succeeding block (Nofer et al., 2017).

Transactions are often inserted into blocks and hashed. The process starts hashing each transaction. Once this is done, the hashes obtained are hashed again, this time in pairs, until a final single hash is obtained. This single hash is called the Merkle Root and is contained in the head of the block, while the rest of the transaction information is contained in the body of the block, as can be seen in *Figure 1.5* (Liu et al., 2020). This procedure is called Merkle Tree and consists in hashing and combining a data structure, until there is only one hash to represent the entire structure (Lesavre et al., 2020). The Merkle Tree relieves nodes from significant storage load, since without transaction details, the space occupied by blockchain data is significantly reduced. Although nodes that hold all blockchain data, including transaction details, are still present, such nodes, called heavy nodes, are a minority. Generally, the Merkle Tree is used to recover disk space occupied by old spent transactions. On the other hand, the Merkle Tree also allows to prove the existence of a transaction in the blockchain, without including in the proof the rest of the transactions in the block (Bernal Bernabe et al., 2019).

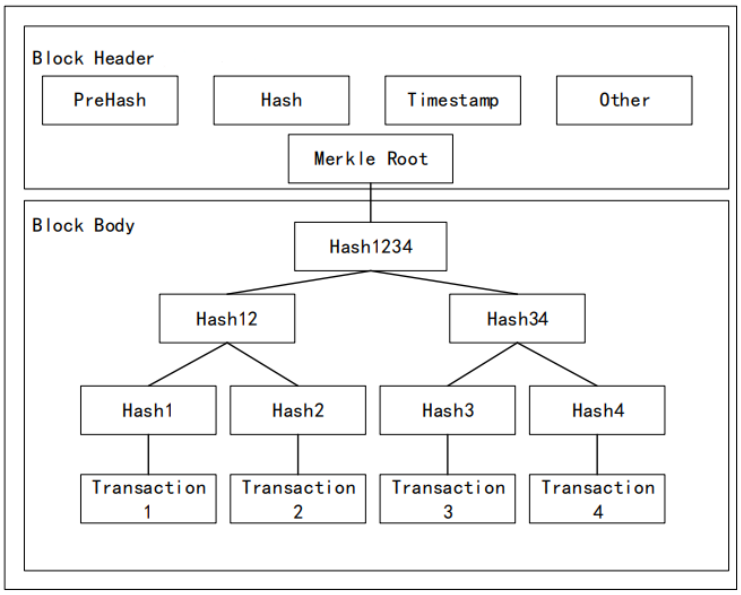

Figure 1. 5 Overview of the Merkle Tree in a block (Liu et al., 2020)

According to the articles of Bochem & Leiding (2021) and Nofer et al. (2017) a nonce is a number used only once in PoW systems to vary the input to a cryptographic hash function to obtain an output hash that fulfills certain arbitrary conditions, established in a decentralized manner. Taking a step back, it was said that in order to validate the transactions and the block that contains them and add it to the chain, it is necessary that most of the nodes of the network agree through a consensus mechanism. Consensus mechanism is a set of rules and procedures used to achieve agreement on which blockchain transactions are valid and which are not. They protect networks from malicious behavior and external attacks. In the most popular blockchain platforms, such as Bitcoin or Ethereum, the consensus mechanism is PoW. In this approach, called mining[9], different nodes, the miners, work to solve a computational problem of variable difficulty that once solved permits to validate the block. It is so clear that new transactions are not automatically added to the ledger. Rather, they are stored in a block for a certain time to the ledger. This time is one of the factors that influences the difficulty of the computation problem. Indeed, the difficulty is defined by an algorithm that verifies how much computational power is trying to solve the problem at a given moment and compares it with the time needed to solve previous similar problems. If there are several nodes trying to solve a problem the computational power available is high and the average resolution time will be low. In this case the algorithm increases the difficulty, to keep the time between two consecutive blocks as constant as possible. This computational problem consists precisely in finding a nonce adequate to obtain a certain target hash, which for example starts with a certain number of zeros, established a priori from the aforementioned algorithm. Due to the characteristics of the hash function, it is not possible to predict what nonce will be required to obtain a hash with the desired characteristics. However, it is possible to continuously vary the nonce in a random way, until the desired hash is found. Note that increasing the number of leading zeros increases the difficulty in finding the nonce and so the time required. Discovering the hash with the required zeros is actually the proof that work has been done. Indeed, finding the right nonce for the target hash is a complex operation, which requires several attempts and consequently a lot of time and resources (i.e., equipment, electricity, etc.) (MacKenzie, 2019). For this reason, the miner who solves the problem first, not only gets the right to validate the block and add it to the blockchain, but also a reward and sometimes also the transaction fees contained in the block itself, which play the role of an incentive.

An alternative consensus mechanism is Proof-of-Stake (PoS) (Houtan et al., 2020). The PoS methodology is increasing its popularity due to the high energy consumption required by PoW. For example, Ethereum's developers have said that they intend to shift to it soon (MacKenzie, 2019). PoS replaces PoW's competition by randomly selecting users to append to the blockchain and earn the associate reward. In this case, mining does not require specialized hardware or solving

[9] Mining is the process used by blockchains to verify new transactions. Vast decentralized computer networks are involved in the process. They verify and protect the blockchains and virtual ledgers that document the transactions that have taken place.

complicated problems. Everyone can solve the problem easily, reducing energy expenditure to negligible levels. However, this comes with some drawbacks, the lack of cost coupled with the benefit of the block reward implies that a validator will always update the ledger whenever given the opportunity, even if the update generates disagreement (Saleh, 2021). Both MacKenzie (2019) and Saleh (2021) show some methods by which to solve this problem, however they also highlight that there are still doubts as to whether the PoS is functional and safe.

Blockchain technology, as described above, offers many benefits. First of all, it solves the problem of Single Point of Failure, since the network remains active even in case of failure of particular nodes. Users don't have to evaluate the reliability of the intermediary or other participants in the network. This increases confidence in the system as a whole. Furthermore, in centralized systems personal data passes and is collected by intermediaries. This causes a loss of control of the data by the user, not to mention that third parties could be subjected to attacks, which result in a violation of the security and privacy of the participants. Thanks to the blockchain, always considering the public permissionless case, intermediaries become obsolete and unnecessary for the functioning of the system, increasing user's security (Nofer et al., 2017). Furthermore, blockchain improves data quality, guarantees their integrity, and promotes their sharing. Additionally, due to its functioning and its block structure, it can be said that the technology offers better control over transactions and their secure storage (Ali et al., 2021).

Blockchain can bring potential benefits to governments, such as minimizing human errors, reducing complexity, providing information anonymity, and improving the voting system. This, combined with all the other general benefits, can reduce disputes and intermediaries in transactions, improve justice, and lessen cybercrimes and corruption (Bernal Bernabe et al., 2019). Other sectors on which blockchain has a strong impact are, as mentioned above, finance, healthcare, and manufacturing. For example, it is possible to identify some specific benefits, such as the possibility of adopting new business models (Ali et al., 2021) or facilitate the introduction and use of Internet of Things systems, in the case of manufacturing (Venkatraman & Parvin, 2022).

Nofer et al., 2017 highlights that there are also risks associated with blockchain technology and high potential for improvements. Among the main problems, difficulties in scalability emerge. They have several causes, including delays in the confirmation of the transaction, data retention, and communication failures. Additionally, some blockchains use consensus mechanisms, which require each participating node to verify the transaction. This limits the number of transactions a blockchain platform can process. Other problems related to the consensus mechanism, already mentioned, are inefficiencies in permissionless systems and high consumption of energy if the PoW mechanism is used. Lastly, key management can be a problem for users of some blockchains. There is a risk that the private key is not well guarded and is lost or stolen. In the first case it becomes almost impossible to recover the key, effectively losing everything related to it. In the second case, the data or cryptocurrencies owned can be improperly obtained from malicious people.

1.2.3 Blockchain & Digital Identity

As seen in paragraph 1.1, when it comes to identity we are still in a highly centralized world. However, the centralized model is facing more and more challenges, with more and more difficulties. From the constant increase in data breaches and identity theft, which lead to reputation damage for entities and a loss of privacy for the people involved, to the increasingly insistent requests for control and security from users. Several research projects for an alternative identity management system have been launched, to expand the trustworthiness and reach of digital identity (Dunphy & Petitcolas, 2018).

Blockchain technology is particularly suitable for digital identity systems, and SSI ones in particular, even though it is not a prerequisite of such models. In particular, decentralization property makes it possible to return control to users, since information is saved on the distributed ledger, and is not in the possession of a central authority anymore. On the contrary, in this way the identity is in the possession of the user only and will not depend on any IdP (Dunphy & Petitcolas, 2018). Obviously, as already mentioned above referring to more general cases, the blockchain allows to solve the problem of the Single Point of Failure (Nofer et al., 2017). This is a problem that plagues, among others, Centralized Digital Identity models and should be overcome thanks to the blockchain (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020).

The blockchain can also improve the problem of Inaccessibility to digital identity. Traditional identification systems often involve complicated bureaucratic procedures, limited access and in the case of developing countries difficulties due to the lack of proximity of the institutions. Therefore, people are unable to have their own identity, as reported by different reports, including McKinsey (2019) and World Bank Group (2017). Blockchain-based identity can gain momentum, as most people without access to digital identity have access to mobile phones instead. Through blockchain solutions, users can, for example, simply use an app for authentication. Consequently, it is possible to reach and give a digital identity to more citizens (Jacobovitz, 2016).

To date, users have multiple digital identities and have to manage several related usernames and passwords, a process that often proves complicated and time-consuming. The option to have a unified identity through which log into any service facilitates management at the expense of security, so it is not viable (Digital Identity Observatory of the Politecnico di Milano, 2020). Furthermore, in many cases there is a limited association between digital and physical identities, which facilitates the creation of fraudulent identities (McKinsey Report, 2019). However, the blockchain, thanks to its structure and the use of cryptography, could stem these problems. As long as the system is adopted by everyone, information on digital identity would be possessed only by

the owner (for example on their smartphone) and would allow access to any service. Therefore, blockchain has the potential to introduce unique digital identities (Fridgen et al., 2018).

For what has been said so far, it is clear that blockchain technology lends itself to identity management. This explains the proliferation of different identity projects that use the blockchain, as evidenced, for example, by Jacobovitz (2016) and Liu et al. (2021). In general, it is possible to distinguish blockchain-based identity projects into two categories:

- **Decentralized Trusted Identity**: identity that is provided by a centralized service that performs identity proofing of users based upon existing trusted credentials, and records identity attestations on the blockchain for later validation by third parties. Proving of the user's identity relies on a general trusted method or recognized documents, such as national identity or passport. An example of Decentralized Trusted Identity models is IDchainZ (Dunphy & Petitcolas, 2018).
- **Self-Sovereign Identity**: identity that is owned, controlled, and managed by the user, without the need to rely on any external authority and without the possibility that this identity can be taken away. It can be enabled by blockchain that facilitates the recording and exchange of identity attributes, and the propagation of trust among participating entities. Some examples of SSI, using the blockchain technology, are Sovrin and uPort (Dunphy & Petitcolas, 2018). This topic will be resumed and explored deeply in the next section.

1.3 SELF-SOVEREIGN IDENTITY

1.3.1 Definition

The term Self-Sovereign Identity was first used in 2016 in a post by Christopher Allen, where he defines SSI as a model in which "*the user must be central to the administration of identity. That requires not just the interoperability of a user's identity across multiple locations, with the user's consent, but also true user control of that digital identity, creating user autonomy. To accomplish this, a self-sovereign identity must be transportable; it can't be locked down to one site or locale.*" It continues adding that "*A self-sovereign identity must also allow ordinary users to make claims, which could include personally identifying information or facts about personal capability or group membership. It can even contain information about the user that was asserted by other persons or groups*." And concludes stating that "*A self-sovereign identity must defend against financial and*

other losses, prevent human rights abuses by the powerful, and support the rights of the individual to be oneself and to freely associate."[10]

Allen's definition is very broad and highlights the main aspects of SSI in general terms. Shortly after, the Sovrin Foundation in its White Paper (2016) proposed a definition that describes characteristics similar to those already stated. However, a part that describes the SSI model in a more pragmatic and practical way was added: "*the best way to think of self-sovereign identity is as a digital record or container of identity transactions that you control. You can add more data to it yourself or ask others to do so. You can reveal some or all of it some of the time or all of the time. You can record your consent to share data with others, and easily facilitate that sharing. It is persistent and not reliant on any single third party. Claims made about you in identity transactions can be self-asserted or asserted by a third party whose authenticity can be independently verified by a relying party*".

Also other definitions such as that of Satybaldy et al. (2019), Pöhn et al. (2021), and Giannopoulou & Wang (2021) focus on aspects such as the individuals' ownership and control of their own identity, the independence from any centralized authority, and security. Proving that these aspects are recognized by the literature as founding and key features of the SSI model.

Der et al. (2017) affirms that Self-Sovereign identities give the person more control over their digital identity. However, they also highlight the behavioral shift required to adopt such a model. Indeed, "*the person now is responsible for the measures taken to establish and maintain both privacy and trustworthiness. Since the digital identities are not issued by third parties, trustworthiness is achieved by the person obtaining evidence for the correctness of the information contained in the digital identity from third parties*".

Also Čučko & Turkanović (2021) propose their vision on the concept of Self-Sovereign Identity and once again there is a focus on some different aspects, not previously mentioned. They describe SSI as a decentralized identity approach that enables entities, not just individuals, to fully control their digital identity without relying on any external authority, eliminating the Single Point of Failure problem. SSI presents a paradigm shift in power and control, from IdPs and SPs to users, who must be central to the administration of identity and information flow during digital interactions.

Although the term Self-Sovereign Identity is still only loosely defined, it is possible to delineate some key properties of the concept. SSI is essentially an identity management system that allows people to fully own and manage their digital identity. However, several academic works, including the Blockchain and Distributed Ledger Observatory of Politecnico di Milano Report (2020), Mühle et al. (2018) and Soltani et al. (2021), agree on the fact that an identity management model can be

[10] Available at: http://www.lifewithalacrity.com/2016/04/the-path-to-self-soverereign-identity.html (accessed August 2022)

considered SSI, if it respects the ten principles defined by Allen and illustrated in detail in the next paragraph.

1.3.2 SSI Principles

Using Blockchain and Distributed Ledger Observatory of Politecnico di Milano (2020) Dib & Toumi (2020), Soltani et al. (2021) to inform the discussion, Allen's ten SSI principles, which an SSI model should possess, are:

1) **Existence**: users must have an independent existence. An SSI must therefore be based on a physical identity and cannot be exclusively digital.
2) **Control**: users must control their identities. They should have the liberty to manage, update or even hide their identities in any way desired as they have complete authority over their identity data. Note that there is a distinction between control and data ownership. A user can have control over an identity claim issued to him, but he may not be the owner of that claim. For example, the driver's license is issued and can be revoked by the government. However, the user should be able to control and share their driving license information on their own terms.
3) **Access**: users should be able to access their identities and all related data easily and directly. There must not be any personal data hidden from its owner. Note that this does not imply that any identity holder can change all the claims associated with their identity, but rather it means that they should be aware of such actions. Furthermore, this principle applies to each entity with respect to its own identity, and not to that of others.
4) **Transparency**: the way an identity system functions, is managed, and is updated must be publicly available and reasonably comprehensible. The used algorithms must be opensource, free, and as independent as possible from any particular organization or architecture, preventing lock-in effects.
5) **Persistence**: an identity must be long-lasting, from birth to death. A claim associated with that identity can be updated or removed, but the identity must be long lived. Identity can only be removed by its owner. This last concept is known as the *right to be forgotten*. It means that individuals have the civil right to request a third party to remove their personal information from the internet (Rosen, 2011).
6) **Portability**: an identity, as well as credentials and attestations, must be transportable by their owner. This is a necessary process for the longevity of the identity data. A digital identity cannot be restricted to a single platform or third-party entity.
7) **Interoperability**: an identity must be as widely usable as possible. A true SSI is globally adoptable and must not only be limited to certain activities and application fields. Therefore,

an identity should be usable by as many entities as possible, regardless of boundaries, jurisdictions, and architectures. This policy supports the availability and durability of the identity.

8) **Consent**: Users must freely agree on how their identity attributes and data are exploited. Identity's information must not be shared without having a consent from the user. Only users should be able to share their data.
9) **Minimization**: disclosing identity attributes must be minimized. Only the necessary piece of information must be shared. For example, when accessing a service, it is necessary to prove that you are of age, instead, to reveal the full date of birth, it is only revealed that you are actually of age. This is achieved using Zero Knowledge Proof, selective disclosure, range proofs, and other privacy-preserving techniques to ensure only the necessary data are disclosed.
10) **Protection**: the rights of identity holders must be always preserved. In cases where there is a conflict between the identity holder and the network, the network should still preserve the rights of the identity holder, even at its own expense. Furthermore, the SSI architecture should be decentralized to avoid possible censorship and monopolies.

1.3.3 Potential benefits

The major potential benefits offered by SSI are listed below:

- In the case of SSI there are no longer central authorities and related databases. This not only permits to solve all the limitations of the Centralized Identity model, described in section 1.1.3.1, but above all to give back control of their own data to the users themselves (Soltani et al., 2021).
- The SSI model guarantees a unique, long-lasting, easily, and widely usable digital identity (Soltani et al., 2021). Having a single digital identity, rather than multiple, facilitates management and allows for a unique reference to be used online and not. Furthermore, it eliminates the necessity of collecting documentation that has already been collected elsewhere. In this way the processes involving the use of identity are more efficient and less time consuming, both for users and for SPs. (Laatikainen et al., 2021).
- SSI's principles of minimization and consent ensure greater privacy and security for users. (Soltani et al., 2021). They also safeguard users against tampering, data theft, and unauthorized monitoring of information. A further benefit remaining on the subject is due to the recent introduction in Europe of the GDPR, which applies to all entities operating in the continent. Indeed, implementing solutions that respect privacy and protect user safety

is increasingly important for organizations that need to be compliant with new regulations. For companies, SSI can offer new ways to fulfil their duties. (Der et al.,2017).

In addition to these general benefits, the SSI can bring specific advantages for different application areas. Between these, Janssen et al. (2020) recommends the use of SSI to make the operation of Big Data Algorithmic Systems (BDAS) more trustworthy. Körner et al. (2022) argue that the use of SSI offers the opportunity to exchange data and even to enforce the business logic in electricity systems. Furthermore, according to the authors, SSI contributes to enhance security and ensures a more integrated management and control of systemic risks. Hasan et al. (2020) propose using SSI and blockchain for digital medical passports and immunity certificates for COVID-19 test-takers, which could reduce the response time of the medical facilities, alleviate the spread of false information by using immutable trusted blockchain, and curb the spread of the disease. Houtan et al. (2020) broadens the field, describing some benefits for the healthcare sector. According to Cocco et al. (2021) the SSI model can be used to guarantee the quality of the products marketed and the compliance of the several supply chain's nodes to standards and technical regulations. In this case, SSI can facilitate the transition from the traditional system, still based on paper or at least on basic IT tools, as well as increasing the efficiency of the system.

1.3.4 Related technical elements

Although the concept of Self-Sovereign Identity is still in its infancy, and there are no common technical standards, it is still possible to identify recurring technical elements. These are described individually in each subsection of this paragraph.

Before getting into the specifics of each individual topic, it is useful to make a quick summary of the main players in the SSI model, quoting the descriptions given by Blockchain and Distributed Ledger Observatory of Politecnico di Milano (2020):

- **User**: the person in possession of the identity.
- **Issuer**: trusted third party that issues claims with certified information associated with the user's identity.
- **Verifier**: entity interested in verifying information on the user's identity.

1.3.4.1 DID & DID document

SSI allows users to generate and manage unique **Decentralized Identifiers (DID)** independent of any third party (Čučko & Turkanović, 2021). Traditional identifiers (i.e., name, surname, email, username, etc.) are provided by the authority issuing the identity, instead DIDs are created by the

entities themselves. DIDs are therefore independent of any centralized registry, IdP, or a certificate authority. Thus, it is possible to say that DIDs are identifiers that possess the characteristics of traditional identifiers plus other innovative properties, which make them particularly suitable for application in the SSI environment (Dib & Toumi, 2020). More in detail, one or more DIDs can be created independently by each user, even offline. Furthermore, each individual can prove to everyone that they actually possess a certain DID without going through a trusted third party (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020).

A DID can be easily created using asymmetric encryption. This technique involves associating two keys, one public and one private, to the DID. The public key can be shared publicly with other entities, while the latter must be kept for the DID owner. The latter is basically a random number large enough to be considered unique. Using this private key, the entity can digitally prove the ownership of a DID (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020). A DID is a permanent identifier in the sense that it never needs to change. It gives an entity a lifetime encrypted private channel with another entity. Indeed, it is possible to use it not only for authentication, but also to exchange messages and Verifiable Credentials. Once a DID is created, its public part should be registered on a distributed ledger so that the actors involved in the relationship can look that DID up. Note that when blockchain is used for identity management, no personal data should ever be put on the ledger. This is because distributed ledgers are immutable, meaning that anything is put on the ledger cannot be modified or eliminated. For this, only the issuer's public DID is stored on the blockchain (Dib & Toumi, 2020).

It is possible to distinguish three different types of DID:

1) **Anywise DID** or **Public DID**: a DID intended for use with an unknowable number of parties. It needs to be resolvable by anyone, without establishing a relationship (Soltani et al., 2021).
2) **Pairwise DID**: a DID intended to be known only by its holder and one other entity, such as a SP. Individuals can have multiple pairwise unique DIDs that cannot be correlated without their permission. In this case DID needs to be resolvable only by the parties in the peer-to-peer relationship (Mohammadzadeh et al., 2021).
3) **N-wise DID**: a DID intended to be known by exactly N number of entities including its subject. It has the same characteristics as the Pairwise DID, which is a particular case of an N-wise DID with N = 2 (Soltani et al., 2021).

Every DID has the following format as defined by the W3C: <Scheme>:<Method>:<Method Specific Identifier>. W3C group has developed this recommendation format for decentralized identifiers to standardize them. The first part of each DID is the DID scheme, followed by the DID method. The third part consists of an identifier in the context of a DID method. The DID method defines the specific methods a DID scheme can be implemented on a particular DLT or network. This includes creating, reading, updating, and deleting (CRUD) operations (Soltani et al., 2021).

The string, which follows the W3C format described above, associates a DID with a **DID Document (DDO)** to ensure secure and reliable interactions among subjects. When a user acknowledges a claim from an issuer, the corresponding DDO is generated (Mohammadzadeh et al., 2021). A DDO is JSON-LD[11] document giving additional information related to the DID. More in details, a DDO, as defined by the W3C, includes the following components: DID itself; cryptographic materials, including public keys and authentication mechanism; cryptographic protocols to interact with the DID subject; the list of DID endpoints; auditing timestamps; a JSON-LD signature to verify the document integrity. A DDO therefore contains the information necessary to establish a communication channel with the owner of the DID (Dib & Toumi, 2020). The need to create a DDO arises because the DID itself does not contain any information. However, it may be useful for the owner of a DID to establish some additional information and perform specific operations. For example, specify a communication channel with which it can be contacted. Delegate someone to sign for him and if necessary, revoke this possibility. Change the public / private key while keeping the same DID. Or even specify some recovery policy for a compromised key. The DDO can contain all this information and allows the correct maintenance of the identity. Also the DDO, similarly to the DID, is controlled directly by the user through the private key. The operation of recovering the DDO connected to a certain DID is called resolution (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020).

Another function of the DIDs is the **DID Auth** protocol. It allows an identity owner to use their client application, such as their mobile device or browser, to demonstrate to an SP that they have effective control of a DID. The DID Auth protocol is based on a challenge-response cycle customized according to the situation. This protocol can replace the use of traditional systems, such as username and password, as a form of authentication and allows for secure and authenticated communication between an identity owner and an SP (Soltani et al., 2021).

Lastly, there is the **DID Comm**, which is a DID-based protocol through which two or more entities can communicate asynchronously, privately, and securely. The protocol is based on DID and supports mutual authentication between the parties (Soltani et al., 2021). DID Comm supports both centralized and decentralized communication models. All the information necessary to establish a DID Comm channel is present in the DDOs of the parties that want to communicate. In particular, the DID Comm protocol uses public key cryptography, preserving the privacy of communications as no one outside the entities involved can know the content and the sender of a message. Each party uses a software agent to process requests and manage keys. Indeed, all interactions take place between the two parties' software agents. An agent can be implemented in a special desktop / mobile application or a web-based application and run within a standard web browser (Mohammadzadeh et al., 2021). The possibility of communicating asynchronously implies the

[11] JSON-LD is a lightweight Linked Data format. Linked Data empowers people that publish and use information on the Web. It is a way to create a network of standards-based, machine-readable data across Web sites. JSON-LD is easy for humans to read and write and it is an ideal data format for programming environments, REST Web services, and unstructured databases (https://json-ld.org/).

presence of two different types of communication, which depend on the availability of the recipient at the time of communication. So, it is possible to distinguish between **direct** and **indirect** messaging. In the first case the sender sends a message to the endpoint specified in the receiver's DDO. Considering the decentralized and ad hoc case, the endpoint is the agent of the recipient, accessible via the Internet. The confidentiality and integrity of the message are guaranteed by encryption and digital signature. To do this, their agents use the other party's public key, which is specified in the DDO. In the second case, that is the indirect one, the two entities are not directly connected, therefore the sender uses an intermediary Relay. In this case the procedure is more complex, indeed, the sender includes his own message already encrypted in another message which will then be encrypted in turn. It then sends it to the Relay in direct communication. The Relay decodes the big message, which includes the original messages, and forwards it to the recipient, again in direct communication. At this point the recipient decrypts the received message and can then read the original message (Mohammadzadeh et al., 2021).

To summarize, Blockchain and Distributed Ledger Observatory of Politecnico di Milano in its report (2020) states that "*Decentralized Identifiers are essentially URLs that associate a subject or entity to a DID Document that allows reliable interactions associated with that subject*". They also identify some key features regarding DIDs:

- **Unique**: globally, but not necessarily unique for the entity it is associated with. A user may therefore have multiple DIDs, but each of them will be unique.
- **Permanent**: does not need to be changed over time.
- **Resolvable**: in a DID Document for which it is possible to trace the metadata on the identifier.
- **Verifiable**: through encryption it is possible to verify its actual ownership by an entity.
- **Decentralized**: registration with a central authority is not required.

1.3.4.2 Verifiable Credentials

Another important building block of SSI, thanks to which most of the value is unlocked, is represented by the usage of Verifiable Credentials (VCs) (Dib & Toumi, 2020). Similarly to DID, VCs are also specifications developed by the W3C Group. A VC is an interoperable data structure suitable for representing cryptographically verifiable and tamper-proof claims (Soltani et al., 2021). It is possible to say that the VCs are enabled by the DIDs. Indeed, thanks to the DIDs it is possible to sign statements, associating them with an identifier. Furthermore, if someone makes a claim about a certain DID, the owner can prove that the claim actually applies to him, simply by proving that he is the controller of the DID the claim refers to. Each statement issued is easily verifiable for this reason it is called a verifiable claim. Each VC consists of a group of verifiable claims (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020).

More specifically, VCs generally consist of:

- **Claim**: that is the set of information and descriptions on the subject to which the digital identity is associated (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020).
- **Proof**: that is the signature that certifies both that the information has been issued by a certain credential issuer and that these data have not been modified (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020).
- **Credential Metadata**: that is the data that contextualize the information of the claim. Among these there are the DID of the subject, the DID of the issuer of the claims, and the DIDs that uniquely identify the credential. In addition, it includes claim expiration conditions, cryptographic signatures, and any revocations (Soltani et al., 2021).

Within the ecosystem of VCs, there are three main entities, based on the role they have. The first is the **user**, an entity that controls one or more VCs. Then there is the **issuer**, an entity that creates new VCs. For example, trusted organizations, such as banks and government agencies, can be credential issuers. The last is the **verifier**, an entity that can verify the VCs it receives. An example of verifier could be an e-commerce website expecting credentials from their customers. In order to use VCs, a verifiable data registry is also needed. It is responsible for mediating the creation and verification of identifiers, keys, verifiable credential schemes and other relevant data (Soltani et al., 2021). In a credential verification process, the holder is requested to present a proof verifying that he fulfills certain requirements. This process is known as **Verifiable Presentation** (VP) and is the base of the use of VCs (De Diego et al., 2021). More specifically, during the authentication process, the holder uses the VCs to generate a VP, which is then transmitted to the verifier. The verifier can then confirm the signatures within the submission to verify the validity of the holder's claim. Notably, this process does not involve the issuer after the initial creation of the VC (Pöhn et al., 2021).

Using Dib & Toumi (2020), Soltani et al. (2021) to inform the discussion, in a decentralized identity framework, the VCs, when transferred, must be understandable and usable by any other system. Otherwise, the VCs should be analyzed manually, which, in addition to limiting their efficiency, will prevent the execution of automated processes and the automatic transfer of identities. To solve this problem, efforts should be made to standardize the schemes that define the structure and content of VCs. In this sense, the direction taken by the W3C Group, by defining the standards for these emerging concepts, can prove to be the most correct one, if effectively understood and applied by everyone. Currently, JSON and some of its specialized versions are the most widely used standard for identity data.

VCs have several advantages over their physical counterpart, such as identity cards, passports, etc. Indeed, VCs not only represent the digital equivalent of physical credentials, but also include

innovative and advanced technologies, like cryptography. This makes them more reliable than physical counterparts. Each holder can, indeed, generate verifiable proofs of their actual possession. Furthermore, both VCs and related demonstrations can be digitally transmitted. Thanks to this it is possible to have faster processes and easier remote authentication (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020). In addition, it is possible to guarantee users greater privacy through the minimal disclosure (or selective disclosure) feature, which is compatible with VCs. Indeed, Zero Knowledge Proof technologies, which will be described in more detail later in the writing, can be used. This allows to reveal only the minimum of information required in an interaction (Sporny et al., 2019). Moreover, VCs let individuals gain greater control over their personal data, indeed, VCs can be shared privately only with intended and authorized parties (Mohammadzadeh et al., 2021).

Like physical credentials, which are usually kept in the wallet, individuals can store their VCs in a so-called digital wallet on their mobile phone, on another edge device, or in the cloud (Sedlmeir et al., 2021). The topic is dealt with in more detail in the next paragraph.

1.3.4.3 Wallets & Key

To enable their usage, Verifiable Credentials should be stored somewhere to make them available when needed. In addition, private keys associated with DIDs must also be securely stored so that they are available for use during proof of ownership. Storing this information is critical to any decentralized identity system. Digital wallets can be used precisely for this purpose. If present in the SSI ecosystem, a digital wallet is a component responsible for storing identity data and cryptographic keys in a secure and privacy-respecting manner (Soltani et al., 2021). More in general, digital wallets can contain various types of information, cryptocurrencies, and crypto assets (Sedlmeir et al., 2021).

There are different forms of digital wallets: mobile phone apps, software, cloud or even hardware wallets (Dib & Toumi, 2020). A user may have more than one wallet. Moreover, a digital wallet, in addition to information storage, can also perform various cryptographic operations such as key generation, signature creation, signature verification, and key backup and recovery (Soltani et al., 2021). Similarly, to all the other components of SSI, it is fundamental that the wallet, and access to it, is controlled only by the entity that owns it. This is not a basic feature that characterizes all wallets, but it should be for those who work with the SSI model (Dib & Toumi, 2020).

As anticipated, there is a strong link between VCs and digital wallets. Among the first to propose this combination there are Reed and Tobin, two key figures of Sovrin, one of the most widespread and established SSI protocols. Their purpose, consistently with the ten principles of the SSI model, is to eliminate any form of centralized control of user claims and make them usable in multiple application areas. More specifically, their proposal is to collect VCs in a digital wallet controlled exclusively by the owner of these credentials. This would make the storage of individuals' identity

information in large government-managed databases or in monetized data silos owned by big tech companies obsolete. In addition, it would no longer be necessary to purchase and transport specialized hardware devices, such as NFC smart cards, encrypted USB wallets or a Google Titan key. On the contrary, a simple smartphone or a desktop alternative are enough to complete the operations related to digital identity. This greatly simplifies the process and facilitates access for more people (Sedlmeir et al., 2021).

The absence of a standard for digital wallets remains an open challenge. Indeed, while SSI supports an open ecosystem, there is value in consistent yet flexible standardization. Above all, because this would also allow greater global recognition by important third parties and also compliance with current regulations (Soltani et al., 2021). For example, Sedlmeir et al. (2021) report that EU COVID digital certificates cannot yet be stored in a standardized wallet along with a wide range of documents, certificates and credentials that can be used to prove a person's identity. In the same paper it is also reported that on the other hand support for VCs and digital wallets is growing, especially in Europe and North America. For example, with the creation of the Verifiable Organizations Network (VON) by Canadian public authorities or the European Self-Sovereign Identity Framework (ESSIF), which uses the European Blockchain Service Infrastructure (EBSI) established by the European Union. There are also private projects such as Trust over IP launched by the Linux Foundation. However, there is also a counterpoint to the support for VCs and digital wallets, namely the resistance of some incumbents. A striking example is described in Sedlmeir et al. (2021). The case tells of internal differences within the W3C regarding the proposal of the W3C Verifiable Claims Task Force to form the W3C Verifiable Claims Working Group, which is able to issue an official recommendation. This is a debate of a political nature, which sees on the opposite side to the proposal, exponents of well-established companies, such as Microsoft, Google, and Mozilla. It is clear that there is still a long way before reaching a common front and much work still needs to be done (Sedlmeir et al., 2021). However, this phase of uncertainty is in a certain sense normal given the recentness and incomplete maturity of the model.

While in traditional identity management models, IdPs are primarily responsible for managing identity data and secret keys and therefore face the responsibilities, risks and technical requirements associated with that activity (Soltani et al., 2021). In other words, it can be said that at the moment people exchange some of their control over their digital identities for a certain comfort, like, for example letting third parties decide on the technology for the secure storage or transmission of data. On the contrary, in the SSI model, this responsibility and the associated risks pass to the users themselves (Der et al., 2017).

There have been numerous instances where users have lost their cryptographic keys, often resulting in the loss of valuable information and irrecoverable funds. Addressing key management requirements in the SSI architecture is a fundamental challenge towards mass adoption of SSI (Soltani et al., 2021). In addition, according to Cheesman (2022), such a key management system

would likely require a demanding set of digital security skills and safety nets. This raises several concerns in the author, especially related to the risk that this type of exacting system would disproportionately exclude already marginalized and disadvantaged groups, like elderly people and disabled. This challenge can be addressed in different ways, for example relying on decentralized key custodians (Soltani et al., 2020), but also using a portable digital wallet (Bandara et al., 2021). In the first case, a decentralized system capable of performing key management operations including key generation, backup, and recovery is proposed. This system uses a wallet, capable of executing the entire model and blockchain technology. More specifically, the proposed key custodian's system relies on a Threshold secret sharing protocol to split secret keys into multiple key share bundles which are shared with a set of key escrow providers. The secret keys are successfully recovered when a sufficient number of key escrow providers reproduce their key share bundle. The blockchain framework is used to register the key escrow providers, to broadcast their capabilities, and to ensure encrypted communication. In the second case, using multiple devices such as a phone and a personal computer at the same time could be a solution. Indeed, whenever a device loss occurs, the user can use one of his other devices to revoke or rebuild his identity. However, although rare, a user could lose all the various devices, thus not having solved the problem. There are also more advanced solutions that involve creating a backup copy of the user's identity data on a cloud, managed by a third party. This backup data can be encrypted using a password known only by the user or using the individual's biometric data. While this may solve the problem, the safety of such approaches is not complete. Indeed, it is not recommended to store sensitive data within a third-party cloud storage even when the data is encrypted or hashed. Storing encrypted data in a certain place that is not directly controlled means giving infinite time to break the encryption and obtain the original data. And this is very dangerous when the data refers to personal attributes (Dib & Toumi, 2020). Another possibility is to appoint delegates during identity creation, who can vote to replace the public / private key pair in case it is lost. The delegates and the dynamics of the voting are established by the identity holder. This is the recovery system used by uPort, now known as Veramo, a very popular SSI protocol. Also in this case, there are problems such as the risk of collusion on the part of the delegates. Key recovery is a necessity for a working SSI system, since key losses are inevitable (Mühle et al., 2018). Even if there is no prevailing method and there are some points of uncertainty, the acknowledgment of the literature of this problem is already an important step, as it brings attention towards its resolution.

1.3.4.4 ZKP

Another important cryptographic technique that can be employed in the context of SSI is Zero Knowledge Proof (ZKP). In particular, ZKP focuses on one of the ten principles of the SSI model, minimization. This principle, briefly recalling what has already been said in Chapter 1.3.2, provides that only the necessary information must be shared (Dib & Toumi, 2020).

ZKPs were first talked about by Goldwasser et al. in 1989. 27 years after the publication, Goldwasser et al. won the Turing award for their effort in ZKP (Soltani et al., 2021). In the last years, ZKP has matured significantly, and to date the topic is followed with strong interest by various people in the environment. ZKP is a cryptographic protocol, which allows a user to convince a verifier of the validity of data with specific properties, without sharing the underlying information (Sedlmeir et al., 2022). The prover thus preserves their privacy, while providing the verifier with sufficient information to be able to validate the veracity of a claim. Therefore, the ZKP protocol improves user privacy, while maintaining the necessary institutional trust for the correct completion of the interaction. The ZKP protocols have multiple application areas, including the validation and the enhancement of privacy of cryptocurrency transactions, which can be considered the most widespread. Multiple examples are present in this field, such as Zcash, ZeroCoin and Monero. ZKPs are also used in Self-Sovereign Identity (SSI) systems, for proof and authentication of identity with respect to privacy, as well as private data mining (Salleras & Daza, 2021).

ZKPs must satisfy three key properties:

1) **Completeness**: if the statement to be proved is true, the prover can always carry out a successful proof. This means that if the input of the ZKP system is true, then it must be accepted as true by the verifier (Bernal Bernabe et al., 2019).
2) **Soundness**: if the statement to be proved is false, the prover cannot convince the verifier that it is true, except for a small probability. So, when the input to the ZKP system is false, it is not possible for the prover to trick the verifier (Bernal Bernabe et al., 2019).
3) **Zero-Knowledge**: the verifier must not learn any information from the proof beyond the fact that the statement is true (or not). In other words, only the prover knows the content of the input to the ZKP system, no extra information is passed on to the verifier or other parties (Salleras & Daza, 2020).

Furthermore, to ensure efficiency and effectiveness, ZKPs must also consider other aspects, such as scalability, interactivity, security and threat modeling, transparency, and quantum security. In addition, they must provide adequate privacy measures to all parties involved. For example, the holder and the issuer should be able to remain anonymous while the confidentiality and non-traceability of claims are still met (Soltani et al., 2021).

There are different types of ZKPs, with different features and functionality. However, in general the ZKP protocol is a type of proof system. Proof systems can be divided into two categories:

1) **Interactive Proof Systems (IPS)**: an interactive ZKP requires the presence of the prover and verifier during the verification process, during which multiple messages are exchanged between the parties. Upon completion of this process, the verifier will either accept or reject the evidence provided by the prover. This system requires an active connection during the exchange between the parties involved (Soltani et al., 2021). However, repeated interactions

are not always a desirable property, as they limit the applicability and efficiency of the system. This is because the applicability is limited to the case in which a prover and verifier are synchronized and simultaneously available (Salleras & Daza, 2020).

2) **Non-Interactive Proof System (NIPS)**: in this case, instead, the simultaneous availability, of both parties, is not required (Soltani et al., 2021). The prover needs only to send a message, the proof, which any verifier can validate off-line. This is also called the public verifier property and it is very useful in the blockchain context. Initially, the NIPS schemes had difficulties in being implemented, due to their impractical computing requirements. However, the advent of Zero-Knowledge Succinct and Noninteractive ARguments of Knowledge (zk-SNARK), one of the most popular ZKPs, changed the landscape. Indeed, this kind of proof can be verified in a few milliseconds (Salleras & Daza, 2020). The importance of zk-SNARK is that they can be used to produce valid proof of the correct execution of a function. This is particularly useful in the blockchain case. Thanks to this method the node that wants to execute a transaction, changing the state of the ledger, can keep the function to be performed, such as a Smart Contract, and the input parameters secret. Indeed, instead of revealing those parameters, the node can load a zk-SNARK to prove that it has performed the calculation correctly and the rest of the participants will trust it. The conciseness makes the proof suitable for being stored in a transaction, and the verification speed allows any other node to efficiently verify the proof (Bernal Bernabe et al., 2019).

As anticipated, interest in ZKP and the related application areas is growing. Among these, there is a particular interest from the SSI ecosystem. Indeed, including the ZKPs in the architecture of an SSI system can guarantee, in addition to the benefits of the protocol, compliance with the minimization principle, since ZKPs limit the amount of data shared to the minimum necessary to achieve a goal. For example, when customers, in a club or shop, are required to prove their age in order to consume or purchase alcoholic beverages. Thanks to the ZKPs, it is no longer necessary to disclose the entire date of birth, since the proof only shows the outcome of the verification, which may for example be suitable or unsuitable (Soltani et al., 2021). More in details, always according to Soltani et al. (2021), ZKP introduces, in the context of Verifiable Credentials, some key capabilities:

- The ability to combine multiple VCs from multiple issuers into a single Verifiable Presentation. This avoids sharing unnecessary VCs or identifiers with the verifier, making it more difficult for it to collude with any of the issuers.
- The ability to selectively disclose the necessary claims with a VC to a verifier, without requiring the need to obtain multiple atomic VCs from the issuers. This allows the holder to only share the information needed by the verifier.
- The ability to produce derived VCs, formatted according to the verifier's data schema, without involving the issuer.

It is therefore clear that ZKPs can be integrated into the architecture of the SSI model and used in combination with VCs. ZKPs help meet SSI's data minimization requirements and prevent unnecessary collection of user data by a verifier.

1.3.4.5 Blockchain

Blockchain technology has already been introduced in Chapter 1.2, in this section the focus is on the relationship between this technology and the SSI model.

As described so far, the SSI model is based on features that do not necessarily require blockchain technology. There are other technologies that allow the creation of an SSI system. However, as will be explained, the points of contact between the blockchain and the SSI model are many and their combination could be the winning one (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020).

To start with, it was said that the SSI model is based on the use of DIDs. An important question, which does not yet have a definitive answer, is that of Key Distribution, which is the verification of the association of a virtual DID with the corresponding real identity. Basically, it has been seen that a verifier can check that a certain claim has been authenticated by a certain DID, however it is also necessary to verify that this DID belongs to a recognized authority deemed reliable in issuing this claim. For instance, a verifier can check that a certain license has actually been signed by a DID, but it must be sure that this DID belongs to the motorization (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020). This is a problem that already exists in classic digital signature systems and online security protocols. To cope with it, various Public Key Infrastructure (PKI) models were created. PKI is a cryptographic protocol that consists of a set of services, tools, processes, and technologies which facilitate the performance of operations based on public key cryptography. The most used PKI certificate template is known as PKI X.509. In this model, a central Certification Authority (CA) creates a X.509 digital certificate. This certificate associates a public key with a particular identity. Together with the CAs, Domain Name Server (DNS) registras and Internet Corporation for Assigned Names and Numbers (ICANN) were created, respectively in 1983 and 1998, in order to facilitate the management and resolution of identifiers and online addresses. The first provides domain name registrations to the general public, while the second is responsible for coordinating the maintenance and procedures of several databases related to the namespaces and numerical spaces of the Internet. In other words, in the traditional case the problem described above is solved by providing the public key, which guarantees the control and a certificate, issued by a CA, which guarantees that this public key is actually owned by a certain entity. Being centralized, the governance model described presents the risk of placing identity data control and decision-making in the hands of a small group of CAs, with the possibility of those authorities misbehaving or becoming victims of security breaches. Following this model, in the case of SSI, the

users should accompany the claim signed by a particular DID with the relative certificate issued by a CA. This certificate guarantees that the DID is owned by a certain recognized entity. However, the solution just described is completely centralized and therefore not compatible with the SSI model. To overcome these limitations, Decentralized Public Key Infrastructures (DPKIs) have been created. DPKI provides a decentralized trust architecture in which no single entity can compromise the security and integrity of the entire system. Differently to traditional PKI, the DPKI does not depend on central CAs. The dependence from these central entities can be eliminated by relying on a decentralized platform as the initial root of trust. There are various DPKI models, however the most widespread and able to overcome most of the limitations of the others is exactly the one based on the blockchain technology (Pennino et al., 2021; Soltani et al., 2021). Through a Blockchain platform, it is possible to implement a DPKI system that is flexible, distributed, and transparent, in order to guarantee the characteristics required by the SSI. In this context, there are no CAs that set and control the key distribution rules, neither that store the keys, but it is the network itself that certifies the identity of the entity that issued the claim, thus enabling the Internet of Trust. It is therefore clear that blockchain technology allows to associate a DID with an identity, without having to resort to a centralized system, thus revealing to be consistent with the principles of the SSI. (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020).

Staying on the DID theme, it is possible to find a further point of contact between the SSI model and blockchain technology. In particular, through the Blockchain technology, it is possible to create a DID Document Ledger that everyone can access, modify without any censorship and in which every change is recorded. If the Blockchain is of the permissionless type, then anyone can create a new DDO, while if it is of the permissioned type in order to create a DDO it is first necessary to carry out an on-boarding procedure. Again, the blockchain is not strictly necessary. However, in order to guarantee the ten principles of the SSI, it is necessary that access to the DDO is not denied by anyone. Furthermore, the creation and updating of DDOs must be carried out transparently and no one can tamper with the recorded information. It is clear that blockchain technology has features capable of responding to these needs. This also makes it particularly suitable for this role of supporting the SSI model, which is the ability to store, allow access, and updating of DDOs (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020).

One of the challenges introduced by the SSI model and VCs system concerns the management of changes and revocations of an issued claim. In the physical case, a document is easily revocable, for example by physically seizing it. This cannot be done for a virtual certificate. It is therefore necessary to have a register for revocations which has the same characteristics as that of the DDO, described above, and which at the same time is constantly updated. Indeed, if the register is not constantly updated, there is a risk that revoked VCs will be misused or that an erroneous claim is exploited to one's advantage in an unethical manner. At the same time, if a public revocation list were created directly online, this would entail very high risks given the sensitive content of the data disclosed and

the possibility of tracing the owner of the revoked data. This problem could be solved using a Blockchain platform, relying also on the ZKPs. With this system, each user can independently generate proof that the credential, of which it is the owner, has not been revoked and this can be shared with anyone, without disclosing any other personal data (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020).

Lastly, blockchain technology is highly compatible and can facilitate the fulfillment of some of the ten principles of the SSI model. The owner of a particular data has full control over it and dictates how such data can be shared with other users within the blockchain domain, thereby satisfying the control and consent properties. Furthermore, a blockchain platform can allow users to implement an immutable standalone program via a smart-contract, which could be leveraged to create a user-controlled IdP coupled with a control mechanism. Thanks to it, users can control access and sharing of their own personal data (Ferdous et al., 2019).

So far, it has been described how blockchain technology facilitates the implementation of the SSI model. On the other hand, SSI can also improve and provide answers to some blockchain problems, including that of strengthening privacy (Bernal Bernabe et al., 2019). More specifically, using the blockchain technology does not require a certified identity, access is allowed using simple pseudonyms. However, there are some situations or applications where a more reliable identification is essential and unavoidable (i.e., custody, legal requirements, etc.). The SSI model is a good option to carry out identification on a blockchain-based platform. Indeed, the claims linked to an identity can be read by a smart contract automatically and without disseminating information. In addition, the user, who must be identified, can provide the identity information signed by the verifying authority and then provide ZKP that the certificate is valid and signed by the authority. The SSI would therefore allow to preserve privacy, maintain decentralization, and make everything more easily readable by programs run on Blockchain platforms (Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020).

All the points of contact described so far allow to affirm that blockchain technology can provide a solid foundation on which to implement an SSI system, while also benefiting from it (Ferdous et al., 2019).

1.3.5 SSI Lifecycle

In this paragraph, it is discussed how the SSI model can be utilized in the different lifecycle activities of identity management. Furthermore, this section can be considered as a sort of summary of the themes introduced so far, since some of the aspects have already been explained previously.

The **registration** procedure can vary significantly depending on the architecture of the SSI model. In general, the registration is the initial step in which the users register themself at the decentralized domain by creating or providing unique values for the identifier (DID) and its corresponding credential. The creation usually takes place through an app for smart devices or a legacy application, where a new key pair is generated. The public key of this pair is used to generate an identifier, which represents the newly created identity by the user. The private key is stored on the personal device in encrypted format along with corresponding public key and the identifier. This is a big difference from the traditional system, as the identity is actually generated by the user himself, without going through a traditional centralized trusted IdP. In addition, also the keys are stored on the user's device and not by centralized entities. Lastly, thanks to the blockchain technology, it is then possible to associate an identifier with an identity, as described in the Chapter 1.3.4.5 (Ferdous et al., 2019). As already mentioned, several times, technologies other than the blockchain can be used in the SSI model. In the specific case, other storage locations can be used for public keys, while the storage of the private key would remain on the user's devices. For example, it is possible to use existing PKIs, as long as they commit to ensuring the fundamental properties of the SSI model (Soltani et al., 2021).

For the part relating to the second phase of the lifecycle, which is **credential management**, the reader is referred to Chapter 1.3.4.2, dedicated to Verifiable Credentials. Indeed, the VCs are the actual credentials used in the SSI model. All information, relating to the issuance and maintenance phase, is described in detail in the mentioned part.

Before being able to access any service, the users must authenticate themselves. **Authentication** is the process of proving the association between an identifier value and the user by providing the corresponding credential value. There could be many ways authentication can be done within a decentralized domain. A substantial difference with traditional methods can be seen in the use of cryptography. Indeed, simplifying, it is possible to authenticate by digitally signing data with the user's private key. The private key is theoretically in the possession of the user only, therefore a data encrypted with this key, and verifiable with the public key in anyone's possession, guarantees that the data comes from the user (Ferdous et al., 2019). As an alternative to data, a random challenge sent by the verifier can be encrypted with the private key. The random challenge is then verified by the latter using the user's public key. These protocols can replace the use of username and password as a form of authentication (Soltani et al., 2021).

Authorization is the process of deciding whether an entity can perform a certain action on a specific resource in a specific domain based on the value of the identifier and other attribute values. It usually takes place in an SP, which checks whether a user can access the requested service, based on certain attribute values. However, it could also take place in an IdP belonging to a decentralized domain (Ferdous et al., 2019). This phase is the one most similar to the traditional model, in the sense that the authorization to access a service is still granted only after verifying the presence of the necessary attributes. However, in the case of the SSI model it is possible for the individuals

thanks to the methodologies described above, such as the ZKPs, to demonstrate compliance with the requests, without disclosing their data. This is a huge step forward in terms of privacy and data security (Soltani et al., 2021).

As regards **identity management**, the methods remain similar to those described in the dedicated paragraph (1.1.2). The main differences lie in the fact that in the SSI model, the attributes are under the control of the users and the CRUD operations can be carried out directly through the platform owned by the individual, such as an App. If the attributes are modified by a third party (for example for revocation purpose), there is still transparency for the user, who will always be aware of what is happening. A peculiarity of the SSI model, in this part of the lifecycle, is the aggregation provisioning, which represents the aggregation operation involving identity attributes by which a specific profile is created for a particular session. Traditional centralized domain generally does not support any attribute aggregation mechanism (Ferdous et al., 2019).

At the end there is the **de-registration** process, which allows an individual to de-register from a decentralized domain by removing the association between the user and the identifier in the corresponding domain (Ferdous et al., 2019). In order to respect the right to be forgotten introduced in some regulations, including the European GDPR (Schlatt et al., 2021), this is a fundamental step in the digital identity lifecycle. The key passage is that at least in theory in the SSI model, the guarantee of the legislature is not needed, indeed as the model is conceived, it should be included a priori (Mühle et al., 2018; Ishmaev, 2020).

1.4 LITERATURE GAP

From the literature review, several gaps were identified. There is a paucity of academic articles focusing on the evaluation criteria to determine whether a decentralized identity solution is SSI or not, as well as a general lack of attention to user experience. However, the biggest gap found is the fact that there is a strong theoretical focus on the concept of SSI, but a little attention to the practical side. This aspect is also highlighted by Cheesman (2022) who states that "*SSI is much discussed but rarely seen in practice*". It is true that there are papers used to describe practical cases, such as Takemiya & Vanieiev (2018), however there is no general research that shows what the trends are in practice. There are also surveys, such as Dib & Toumi (2020) and Panait et al. (2020), which however focus on the analysis of protocols and not on the application cases. Furthermore, even these analyses are often theoretical, while comparisons on practical or user experience-oriented issues are limited.

1.5 RESEARCH QUESTION

Due to the Covid-19 pandemic, digital processes have become increasingly important. The need to avoid physical contact and at the same time to proceed with daily activities has moved many operations, traditionally carried out in person, online. However, this required first of all a reliable and secure digital identity, to authenticate and interact online. The presence, among the others, of critical activities required a very high level of security for these identities, comparable to the classic one. This explains the explosion, especially in some countries, such as Italy, in the adoption of eID and other certified digital identities. In Italy in particular, the total number of SPID digital identities issued has increased from about 6 million pre-pandemics to the current 31.8 million (August 2022)[12]. In the first year of the pandemic, the number of SPID digital identities issued has tripled. (Pöhn et al., 2021). However, the emergency situation has also contributed to underlining the limits of the traditional identity model. It is known that unforeseen and particularly serious events also propagate due to the inefficiency and unpreparedness of the systems in place. In particular, one of the greatest difficulties encountered by the traditional identity system was to balance the need for quick and safe responses, both for the containment of the pandemic and for the continuation of daily activities, with the need to guarantee the privacy standards for users. Often there were situations in which one of these two needs had to be sacrificed in favor of the other. (Bandara et al., 2021).

Also the intensification of migratory flows in recent years, due to the accentuation of some conflicts already underway, especially in Africa and in the Middle East, and the more recent outbreak of the war in Ukraine, has shown further limitations in the traditional identification system. Indeed, at the moment, a series of issues prevent access to formal means of identification for refugees, people fleeing their own country and sometimes even citizens themselves belonging to discriminated minorities or disadvantaged groups. These issues include the loss or damage of documentary evidence, statelessness, and the absence of or exclusion from a national ID system. For example, the passports of many Syrian citizens have been systematically destroyed by Islamic State (ISIS). On the other hand, the need to have an identity must not override the fundamental rights of such people. Indeed, most of the existing digital refugee identification systems show an absence of informed consent. Furthermore, the subjects lack control over how their personal information is collected and used. Additionally, refugees face extensive bureaucratic challenges, attempting to change or update their data. In general, there is very limited transparency around data flows and sharing with third parties, which often lead to the use of humanitarian data for non-humanitarian purposes, such as ad targeting. (Cheesman, 2022).

[12] https://avanzamentodigitale.italia.it/it/progetto/spid

These limits depend on the fact that the identification systems used for the most part are still centralized (Dib & Toumi, 2020). The end users are little more than a number and their interests are often not respected. The answer to these problems, as mentioned above, exists and is the SSI model, which allows people to have reliable digital identities, whose control passes or rather comes back to the individual. Furthermore, the principles on which the model is built allows to overcome the limits seen in the two previous examples and more generally guarantee both an identity valid everywhere and in any case, and privacy and security for users (Mühle et al., 2018). However, as discussed in Chapter 1.4, SSI is a new and still highly theoretical concept, and there are little insights into its practical application. This thesis aims to fill this gap in the literature by analyzing the practical implications of this model, taking a census of the existing use cases and studying them within a specially prepared framework. As such the research question to be addressed is: *how is the landscape of SSI-type systems configured at international level*?

Although interest in the topic has grown significantly in recent years, as have many of the technologies and techniques that can be used with it (i.e., blockchain, ZKPs, etc.), very little research has been conducted to understand how to implement the SSI model in practice. Academic literature focuses primarily on the concept of SSI and its ecosystem at the theoretical level. Articles that illustrate use cases or compare and analyze well-known practical cases are in the minority and are often limited. More specifically, these documents often focus only on protocols and not on the application cases of these protocols, sometimes including without distinction even cases belonging to the Decentralized Trusted Identity model and with an almost always theoretical perspective. The economic model, which one would like to adopt for SSI systems, is hardly ever even considered. No general research that shows what common practices are, were found. However, the intensification of requests for greater privacy and control coming from individuals, combined with the tightening of international regulations increasingly attentive to the interests of consumers, make the transition from traditional to Self-Sovereign models increasingly necessary. In this context, analyzing the current landscape of practical applications of the SSI model can help, both institutions and privates, in identifying the main application trends, the main opportunities and the greatest challenges, with special attention to the cases deemed most successful. Note that cases of identities not for individuals, such as identities for companies or intelligent objects, were excluded from the census. The interest of this research work focuses on the potential of the SSI model for people, which is the field for which it was created.

Based on what has emerged so far, it is possible to identify three further sub-questions to be addressed with this thesis work: *are the ten principles of the SSI model respected in practice? Does the strong link between blockchain and SSI that emerged at the theoretical level also exist at the application level? Is it actually possible to associate any type of data to the SSI?*

Chapter 2: Research Methodology

This chapter describes in detail the different phases of the research project that led to the results, described in Chapter 3. First, the process of Literature Review will be described. It follows a description of the analysis framework, built on the basis of information extracted from the literature review. It ends with a description of how the empirical work and the related analysis were conducted.

2.1 DIGITAL IDENTITY OBSERVATORY

The entire research project was developed in collaboration with the Digital Innovation Observatories of the School of Management of the Politecnico di Milano, which allowed access to all the knowledge accumulated over the years, such as databases and scientific archives, constantly supporting empirical work in its various phases, sharing news, information and providing valuable observations.

In particular, this work was developed with the Digital Identity Observatory, which was created to address issues related to digital identity. It was born on 27th November 2019 as a Working Table, but after understanding its relevance in the digital transition, it became an officially recognized Observatory in 2020. The Digital identity Observatory aims to understand the potential offered by digital identity systems and to contribute to the development of the market in Italy. Furthermore, the Observatory manages a qualified and independent table, in which to encourage cross-industry comparison between market players, through empirical analysis and research, aimed at defining the characteristics and opportunities for the development of digital identity. Among the research objectives of the current year, there are several related to the themes of this thesis. Above all "*identifying the main approaches and applications of decentralized identity models*".

2.2 THEORETICAL REVIEW

To collect the necessary academic material, Scopus was used as the primary source database. Scopus is an online database that collects thousands of academic texts. It allows users to filter them, extracting the main information, through a set of intuitive tools, thus allowing more focused and comprehensive analysis of the literature. The process followed three main steps:

- **Extraction**: in this phase, thanks to the command "Article title, Abstract, Keywords" available on Scopus, it was possible to extract the documents on the topic of interest. Indeed, this command allows users to find all the papers that contain a given keyword in their title, abstract or author's keywords. The keyword used in this stage was "Self-Sovereign Identity" with the addition of some filters, such as English language and Subject Area (like Business, Management, Economic and Social). The output was a CSV file containing different information regarding the papers found, sorted by descending number of citations. In particular, the information downloaded for each document was author, document title, year, source title, volume, issue, pages, citation account, publication stage, abstract, and author keywords.
- **Screening**: in this step the abstracts of all the documents found, about a hundred, were read, classifying the documents into relevant and non-relevant. Those belonging to the first category have been downloaded for a more in-depth analysis.
- **Analysis**: in this stage all papers cataloged as relevant were classified based on the robustness and the reputation of the source. To do this SCImago Journal Rank or SJR indicator (an indicator for measuring the scientific influence of academic reviews) was used. For each journal entered, a quartile is returned indicating the authoritativeness of the source. There are four categories, ranging from best case to falling, and they are Q1, Q2, Q3 and Q4. Subsequently, all the papers found were read and summarized, highlighting the key points and the most interesting aspects for the development of the research.

The output of the Theoretical Review was an Excel file, containing all the documents read and related information, and a Word file, containing the summaries and key passages of the various articles read. Some other papers were extracted from Google Scholar to further enrich the information collected. In addition, reports from practitioner-oriented research centers such as McKinsey, World Economic Forum, World Bank Group, etc. were also included among the readings. The content of the various documents is detailed in Chapter 1.

2.3 ANALYSIS FRAMEWORK

Thanks to the literature review, it was possible to identify the relevant variables for the analysis of the SSI cases. The framework of analysis, used in the empirical work, was constructed precisely using these variables. The identified analysis dimensions are described in *Table 2.1*.

Category	Variable	Description
General Information	Name	The name of the case

	Date	The project date
	Geographical Areas	The country of the project
	Status	The development phase in which the project is (i.e., Announcement, Active, Stopped, etc.)
	Participants	The organizations involved, the type of these organizations (institutions or private companies), their role in the ecosystem
	Model Diffusion	The number of people using the solution
Platform Technologies	SSI Protocol	The SSI protocol adopted
	Ten SSI Principles	Compliance with each principle of the SSI model
	Blockchain	Use of the blockchain and its characteristics (name, public vs private, permissioned vs permissionless)
Integration Technologies	Protocols	Protocols adopted (SAML, OIDC, API & SDK)
Process Technologies	Digital Wallet	Use of Digital Wallet and its characteristics (Desktop, Mobile, Physical)
	Mobile App	Use of Mobile App and its characteristics (use of in-app notifications, PIN, password, QR code, biometrics)
	Others	Use of others type of process technologies (i.e., NFC cards) and their characteristics
Onboarding Procedure	In person recognition	Need for in person recognition
	Online recognition	Need for online recognition and its characteristics
Access / Use Credentials	Linked Document	Use of Social Security Numbers or Physical Document Linked to access
	User ID	Use of User ID to access
	Phone Number	Use of Phone Number to access

	Password	Use of Password to access
	PIN	Use of PIN to access
	OTP	Use of OTP to access
	Biometrics	Use of Biometrics to access
	#MFA	Number of different factors required to access
Data Type	Personal documents	Personal documents (i.e., national ID, passport) are associated to the identity profile
	Banking Information	Banking information (i.e., credit card, bank identity profile, bank certified information) is associated to the identity profile
	Driver's license	Driver's license is associated to the identity profile
	Training certification	Training certifications are associated to the identity profile
	Municipal service card	Municipal service card is associated to the identity profile
	Contact information	Contact information (i.e., email, phone numbers, etc.) is associated to the identity profile
	Biometric Information	Biometric information (i.e., selfie, fingerprint, etc.) is associated to the identity profile
	Health information	Health information is associated to the identity profile
	Graduation certificate	Graduation certificate is associated to the identity profile
	Certified job role	Certified job role is associated to the identity profile

	Gas subsidies	Gas subsidies (i.e., state vouchers) are associated to the identity profile
	Rental history	Rental history is associated to the identity profile
Data Type Certification	Certified	Data associated to the identity profile are certified or linked to physical documents
	Self-declared	Data associated to the identity profile are self-declared
Applications Areas	General Purpose	Solution developed for general purpose or with at least four different specific application areas
	eCommerce & Retail	Solution developed for application in the eCommerce & retail sector
	Finance	Solution developed for application in the finance sector
	Telco	Solution developed for application in the telco sector
	Healthcare	Solution developed for application in the healthcare sector
	Travel & Tourism	Solution developed for application in the travel & tourism sector
	eGov	Solution developed for application in the eGov sector
	Enterprise	Solution developed for application in the enterprise sector
	Humanitarian Scope	Solution developed for humanitarian purposes
	Mobility	Solution developed for application in the mobility sector
	Utility	Solution developed for application in the utility sector

	Gaming	Solution developed for application in the gaming sector
	Education	Solution developed for application in the education sector
	Rental	Solution developed for application in the rental sector
Service Provider	Number	Number of SP in the project
	Type	Type of SP in the project (i.e., public, or private SP)
Economic Sustainability	Business Model	Business Model adopted

Table 2. 1 Summary description of the analysis dimensions considered

A clarification is necessary regarding the dimension of the project status. Indeed, this variable includes several categories. In addition to those self-explanatory such as announcement, active and stopped projects, there are Proof of Concepts (PoC), which are simpler tests, generally carried out by the developers themselves in a close environment. Then there are pilots which, on the other hand, often include a small number of users who act as testers for the solution, often these users are the employees of the companies involved. Lastly, there are projects that are in the rollout phase, but are not yet active.

2.4 EMPIRICAL FRAMEWORK

This paragraph describes in detail the process followed to carry out the empirical work. This process is made up of specific and well-defined phases, each of which is described individually. First came the selection of sources from which the bulk of the necessary information was obtained. Once the sources were selected, the data was extracted. At this point it was necessary to carry out a screening phase, in order to register only relevant and inherent cases with the SSI model. This stage was followed by the data integration phase, which made it possible to collect uniform information on all selected cases. This made it possible to create a list of application cases of the SSI model, on which the analyses were carried out. The different phases of the process are shown in *Figure 2.1* and will be explained in detail in the next paragraphs.

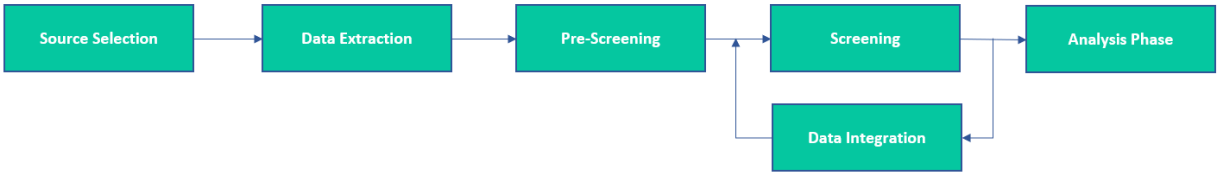

Figure 2. 1 Empirical Framework

2.4.1 Source Selection & Data Extraction

The first step in carrying out the census was to find as many cases as possible related to the world of SSI from secondary sources. This was possible in two different ways. First, while reading the academic articles for the Literature Review process, all use cases mentioned or described in the papers were noted and briefly described. Secondly, a search was carried out on websites specialized in news regarding new technologies. In particular, the following sites were systematically consulted:

- https://www.coindesk.com/
- https://cointelegraph.com/
- https://www.ledgerinsights.com/
- https://bitcoinmagazine.com/

In addition to periodic consultation, research was carried out in the news archives, through the appropriate search function, present on each of these sites. More specifically, the following keywords were inserted in the search engine: "Self-Sovereign Identity", "Self Sovereign Identity", "SSI", "Self-Sovereign", "Self Sovereign", "Decentralized Identity", and "Blockchain Identity". Only the most pertinent cases were extracted from the research in the sites' archives.

The result of this phase was a Word file containing a brief description and some key information of the cases extracted from the two different types of sources. It is clear that this list included cases whose information was not uniform, duplicates and also projects that were not strictly SSI. To improve this, a screening operation was carried out immediately after.

2.4.2 Screening & Data Integration

The screening phase, necessary to solve the problems mentioned above, was divided into two steps:

- **Pre-Screening**: in this stage all irrelevant cases and duplicates found, were eliminated. In addition, short descriptive cards were prepared with some basic information for the remaining cases. These also included some borderline cases, which were further examined in the second phase.
- **Screening**: a detailed analysis of the borderline cases was carried out, based both on information available and other primary sources. These include the official websites of the considered case, its authoritative social network profiles (i.e., LinkedIn), and other sites considered valid for obtaining direct information on the case, such as Medium[13], where the articles are sometimes inserted by the developers of the solution themselves. More in detail, the borderline cases were cases of decentralized identity, not necessarily in line with the SSI principle, and classifiable as Decentralized Trusted Identity. Cases included in this category were then discarded.

The final product was a list that contained only the relevant cases. Based on the dimensions of the analysis framework, defined above, the missing data was integrated using especially primary sources, such as official websites of the case, its authoritative social network profiles (i.e., LinkedIn), and articles on the sites of the protocols or of the organizations involved.

2.4.3 Analyses

The final result was an Excel file containing 51 SSI cases. For each of them all the information described in the previous section has been researched and collected when available. Based on this considerable amount of information, it was possible to conduct the analysis.

First, each dimension of analysis was analyzed, then variables were cross analyzed in order to create a comprehensive representation of the Self-Sovereign Identity ecosystem. Graphical representations were often used to visualize the data in a more intuitive and immediate way. A further specification is necessary, since this is an emerging model, some of the cases found have limited information available, and some dimensions had to be left blank, the most interesting cases were still reported using qualitative observations. The analyses were performed focusing on different aspects:

- **General**: includes the analysis of the general information about the SSI ecosystem. In this way it is possible to know the geographical distribution of cases and the presence of any leading nations. Furthermore, it is possible to verify the progress of the practical application of the model, considering the information obtained from the analysis on the status of the

[13] https://medium.com/

projects and their temporal distribution. It is also possible to identify the type of entities present and understand if there are any particular trends in the composition of the ecosystem. Lastly, considerations, relating to the economic sustainability of the model and its diffusion in the analyzed projects, are made, presenting significant cases.

- **SSI Principles**: includes the analysis of the SSI protocols used and their compliance with the ten principles of the SSI. This allows us to understand how many protocols exist and if some are established. Furthermore, this perspective represents the bridge between the theoretical and the practical part of the SSI system, allowing observations to be made regarding the applicability of the model, and also highlighting the most critical points in the transition from theory to practice.
- **Technological**: includes the analysis of the technological aspects regarding the SSI models. In particular, the link between blockchain and SSI is explored, focusing on the main characteristics of the platforms adopted and the consequences on compliance with the SSI principles. Furthermore, it is possible to identify the most used integration and process technologies, verifying the presence of trends or established practices.
- **User**: includes the analysis of the impact that the model has on users and their experience. This perspective includes both the study on onboarding practices and on identity management through the users' own devices.
- **Data**: includes the analysis on the type of data associated with the identity profile and the relative certification level. This allows to understand which data are most often attributed to the identity and if they are certified or self-declared by the users. This last point enables a reflection on the reliability of the identity and the consequent potential diffusion. Furthermore, it allows to analyze how varied the information collected in practice is, since theoretically any attribute could be associated with the identity. Lastly, this perspective is combined with some others to generate more varied insights that allow for qualitative observations.
- **Application areas**: includes the analysis on the main application areas tackled by the found cases. It also allows to verify if this identity system is limited to certain application areas or if, on the other hand, it can be used in a wide range of application areas. Furthermore, also this perspective is combined with some others to generate more varied insights that allow for qualitative observations. For example, if there are sectors in which government projects are focused, or which are the sectors with the most active cases, and so on.

The results and key messages extracted are described and discussed in Chapter 3.

Chapter 3: Results

The current chapter presents the results obtained after completing the empirical work, described in detail in the previous chapter. Chapter 3 is composed of graphs that illustrate the outcomes of the research carried out, accompanied by a qualitative description that explains them. The results are divided according to the main perspectives, described in Chapter 2.3.3, but also include analyses on several levels, carried out by combining the different dimensions of analysis. In summary, the sample is made up of 51 cases of Self-Sovereign Identity and the main aspects, on which the analysis has focused are:

- General Information.
- SSI Principles.
- Technological aspects.
- User perspective.
- Data.
- Application areas.

Thanks to the information collected during the empirical work, and presented in the following pages, it is possible to answer the Research Question and the related sub-questions.

3.1 GENERAL INFORMATION

3.1.1 Geographical Distribution

The data used for country distribution was collected based on where projects originated and developed, even if some of them may have expanded their businesses to other parts of the world. Instead, the projects that from the beginning planned to carry out the operations in different parts of the world, have been categorized as international projects. To give an idea, this is the case of projects involving airlines operating in several states, such as Known Traveler Digital Identity and Iata Travel Pass. Also cases promoted by international organizations such as the Identity For Good project, which involves the Red Cross, have been placed in this category.

The description begins with *Figure 3.1* which shows the geographic distribution of the 51 cases found. It also highlights the number of different states, in which there is at least one project and the number of active cases, namely those on the market and fully operational.

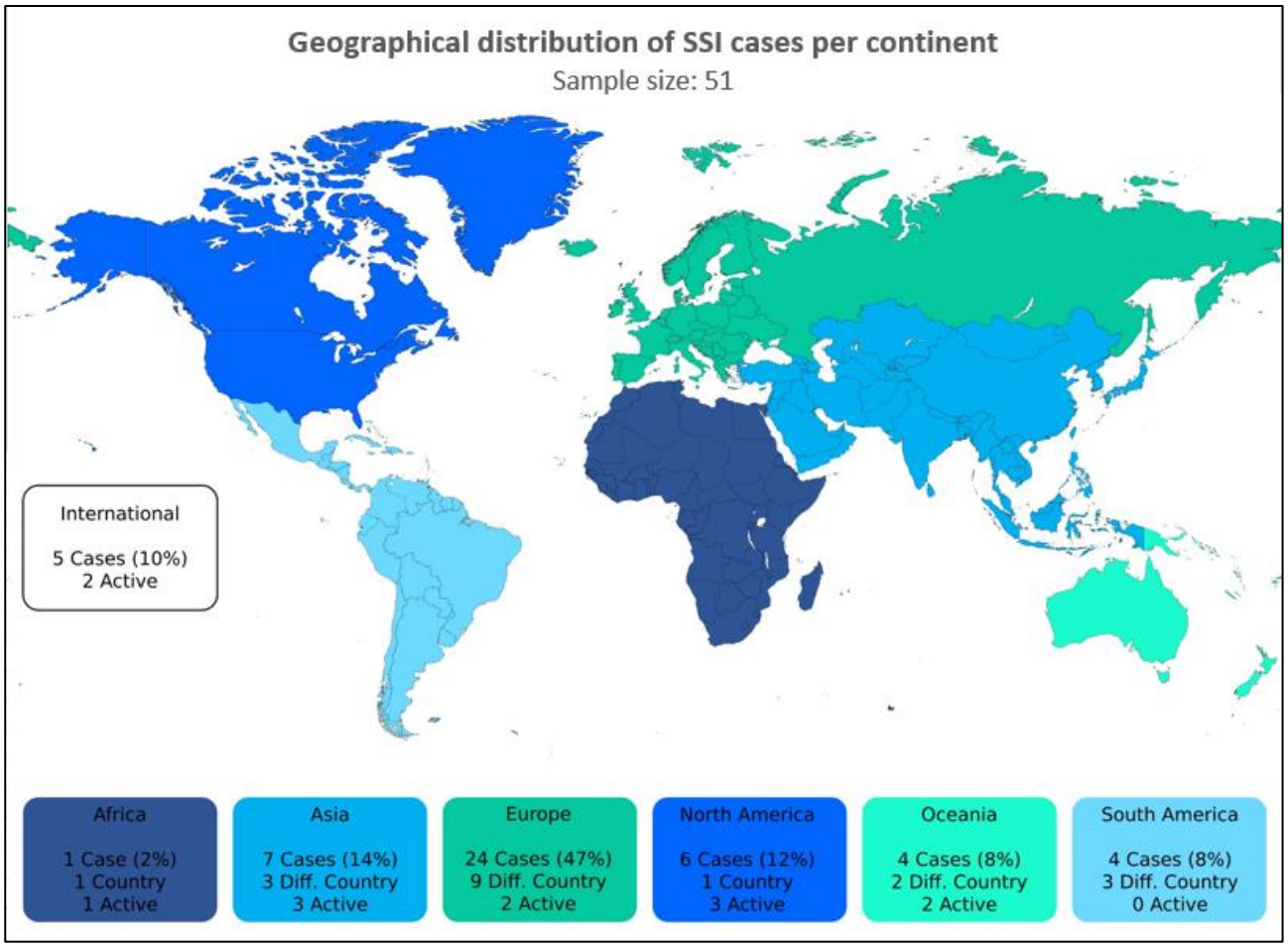

Figure 3. 1 Geographical distribution of SSI cases per continent (the total percentage exceeded the limit due to rounding)

From the analysis it emerges that Europe has a predominance both in terms of the number of cases and of nations involved in the projects. The first three countries for projects (Germany, USA, and South Korea), belong to three different continents and represent about 40% of the total cases (20 out of 51).

Focusing on the results of the analysis of the geographical distribution by continent, two interesting aspects emerge. First, South Asia and especially Africa, which are considered the two most critical continents in terms of lack of Identity (McKinsey Report, 2019; Cheesman, 2022), are effectively scarcely present, at least in terms of SSI. However, the presence of some, albeit few, projects demonstrates that the problem is recognized, and action is being taken to try to solve it. Although, a lot of work is still needed to achieve the goal. The second interesting aspect concerns the number of cases in North America. Indeed, North America is generally considered to be very active in the field of digital identity, for example it is the most represented continent per number of startups operating in the digital identity sector (Di Sarno, 2020). However, it is possible to observe a different trend when it comes to SSI. In particular, as seen in *Figure 3.1*, North America is the third continent by number of cases, “only” accounting for about 12% of the total. Surely a greater interest on the

part of this continent for the SSI could give impetus to the further development of the model. On the other hand, it should also be emphasized how the characteristics of the SSI make it more interesting in other continents, such as Europe, where for some years there has been more and more attention to the protection of privacy and security of the users. Despite these limitations, it should be noted that the United States is the nation with the most active cases, 23% of the total, and with the most general-purpose cases, 29% of the total, compared to having only 12% of the total cases. Recalling the description in Chapter 2.3.2, general purpose projects have no limitations in terms of application areas. These are two rather interesting results. Indeed, although there are few total cases, they are very much in line with the principles of the SSI, especially in favoring interoperability, thus not limiting themselves to serving specific sectors or niches. Furthermore, they are proportionally more likely to achieve the effective functioning of the project.

Finally, it is necessary to make a clarification on the cases found. In fact, the research was carried out in English, which may have made it more difficult to find cases in countries such as China or Russia, for which no projects were found. This is a limitation that can only be overcome if the various works are also promoted in English.

3.1.2 Project Status and Temporal Evolution

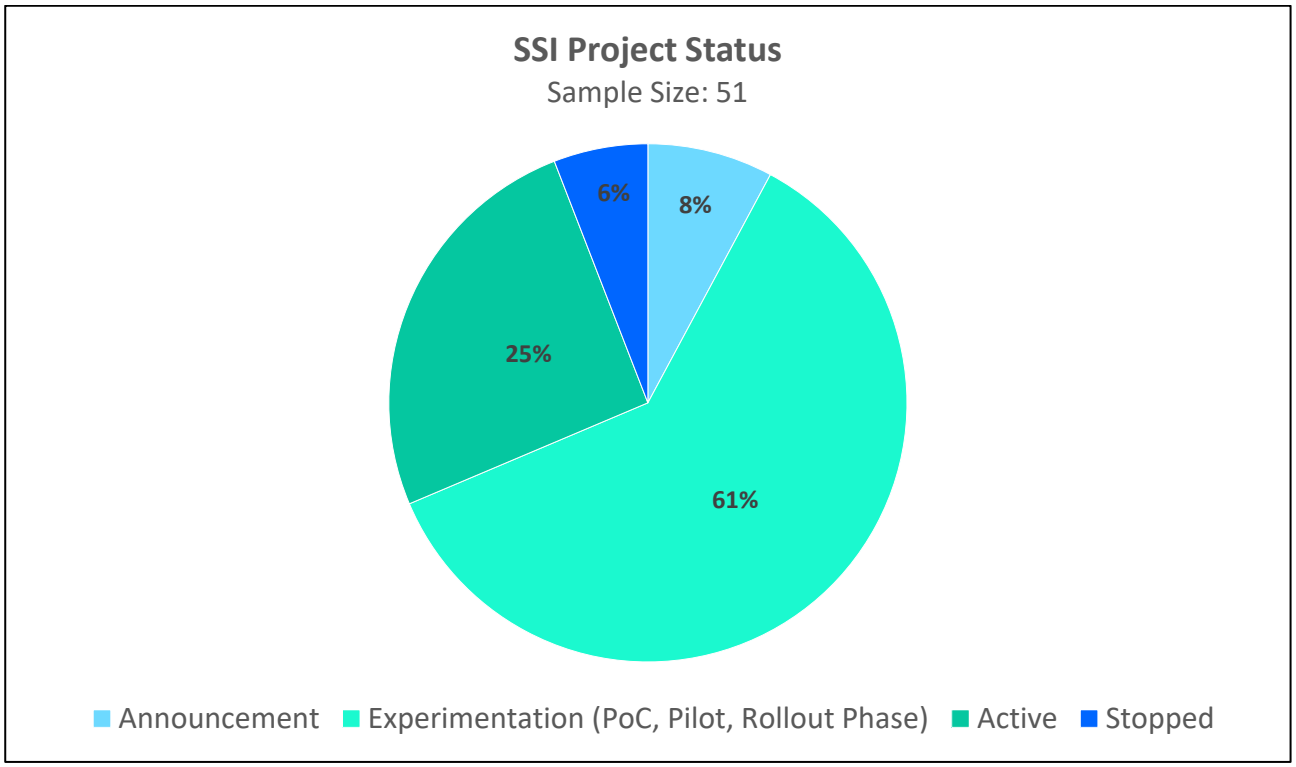

Figure 3. 2 SSI Project Status

From the graph in *Figure 3.2* it can be seen that the tangible projects represent a majority compared to the announcements and the interrupted projects. In a sense this tells that the theoretical or for future research interest, it is turning into concrete and practical interest in the model. However, most of the projects are tests or pilots, while already active projects represent a quarter of the total. This is rather indicative of the embryonic stage in which the practical application of the SSI model still stands. The experimental part is still prevalent over the rest. The state defined as experimentation, includes different types of projects, as explained in the previous chapter. Although, there is a certain difference between these categories, albeit in many cases not too marked, it was decided to combine them in a single voice, to underline the fact that they are practical solutions, effectively tested, but not yet in the final version. This permits to mark the difference with the announcements, which are projects without any prototype, and which may never be developed. At the same time, it is possible to keep a distance from the solutions that are operational and available on the market. As a final remark, it must be said that even projects in the experimental phase could be interrupted or not see a sequel, however their analysis allows to obtain a lot of information regarding the practical functioning of the model.

To better understand the evolution of the model, a combined analysis between the project status and the year was conducted. The main results are shown below in *Figure 3.3*.

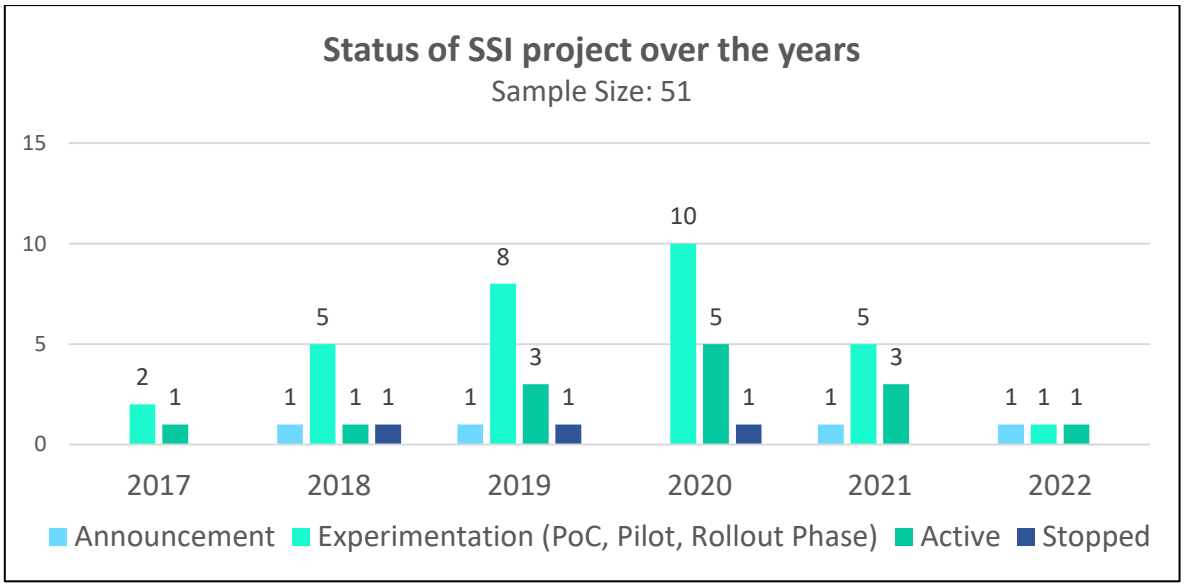

Figure 3. 3 Status of SSI projects over the years, NOTE: the number is greater than that of the sample, because for the stopped cases, the year in which they started and the year in which they stopped are considered separately (if everything took place in more than a year).

It is interesting to note that cases have grown steadily until 2020. Even more interesting is the fact that the percentage of active cases, excluding the first and last year which have a very small sample, has increased over the years. This is a sign of the fact that the proposed solutions are increasingly ready to be distributed on the market. Going into detail, of the 14 active cases only 1 was stopped after 3 years. This is a good sign of the robustness of operational solutions. The other two

interrupted cases are an announcement that never had a sequel due to the blocking of the activities of the two companies involved. While, in the other case, the practical tests stopped after a PoC, due to the need to do further research, differently from what was initially planned.

In addition, there have also been some changes over time. Cases that in the initial data found were categorized as announcements and then became active projects to all effects. From this point of view, an interesting case of evolution is the ONCE project, developed by the German Bundesdruckerei with Jolocom. This is a pilot project that started in 2021, following the completion of a PoC in the previous year. The knowledge accumulated in the first part of the project allowed the designers to further develop the solution. Furthermore, new partners, including some German cities, joined the project. All this has made it possible to release a new solution, which is being tested by a small number of people. The goal is to gather more information and insights from users, in order to be able to further develop and launch a final version soon.

3.1.3 Main Actors

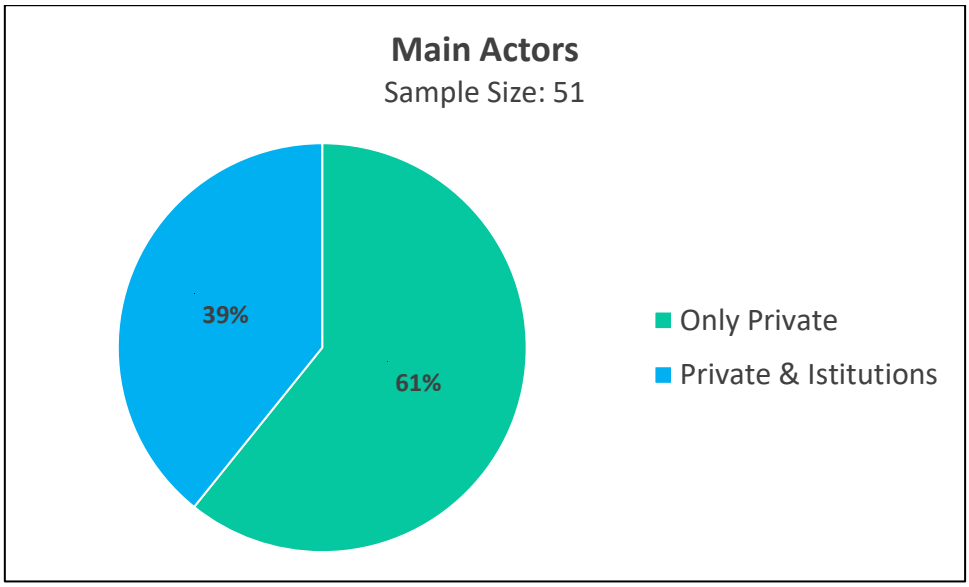

Figure 3. 4 Main Actors

As can be seen in *Figure 3.4* the private sector is driving the model, being represented in all cases, and being present exclusively in most of them. This point can be seen as the natural consequence of the need of the actors in the ecosystem to distance themselves from entities that for years have been the central authorities in terms of identity, which often is precisely the institutions. However, it should still be noted that the presence of the institutions is still significant, a sign that the model has aroused the interest of even the most established entities. However, probably due to lack of specific skills, they prefer to turn to the private sector for the realization of their projects.

Furthermore, as shown in *Figure 3.5*, the difference is even more marked if we consider the active cases, where 85% of cases rely only on the private sector, against the 61% of the overall situation.

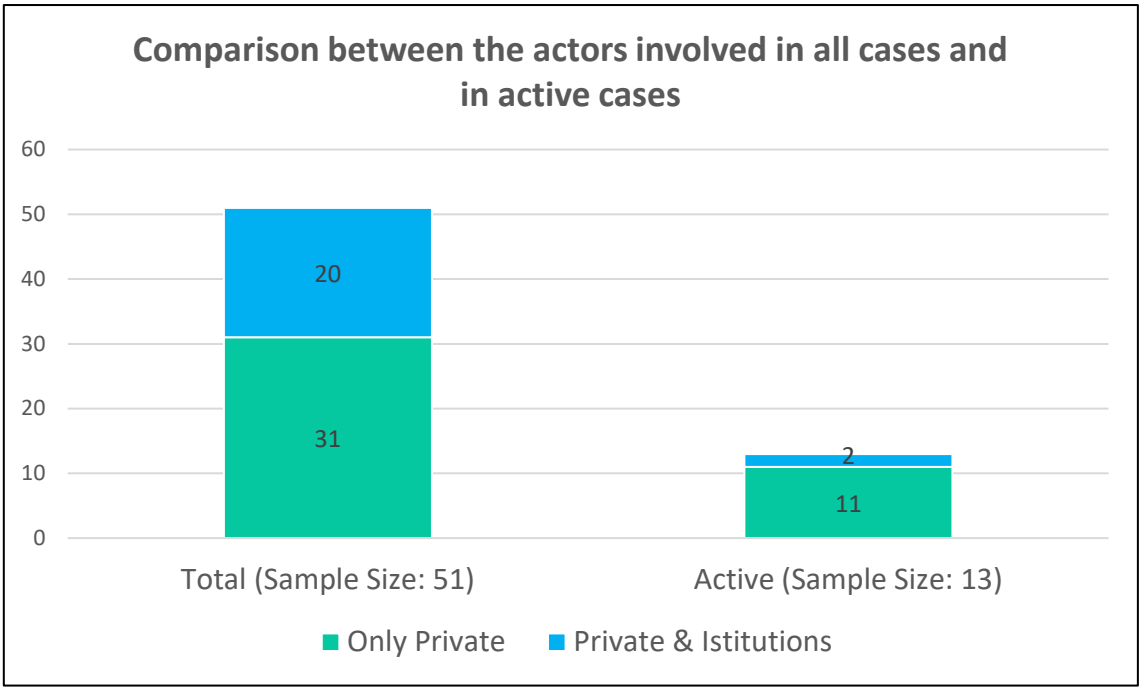

Figure 3. 5 Comparison between the actors involved in all cases and in active cases

This indicates how projects linked with institutions tend to stop before reaching full activity. In a certain sense, it can also be said that the interest of governments is more in evaluating how this model could integrate with the services offered by the public administration, evaluating its applicability and the advantages and disadvantages, rather than immediately pushing for an active project.

In a similar way, an analysis was conducted on the type of Service Provider present in the ecosystem of the various cases. However, there is often a lack of detailed data. This prevented the formalization of significant numerical results. On a general level it can be said that in most cases there are few SPs, and they correspond to the participants of the project. Also in this case the results show a prevalence of the private sector, compared to the public one.

3.1.4 Model Diffusion

Model diffusion is intended in numerical terms, namely the number of users involved in the various projects. It is a mostly qualitative analysis, because the data necessary for this analysis are present only in a few cases, 16% of the total, and are often not even precise values. The rarity of the data inevitably affects the value of the results at a statistical level, in fact, it would make little sense to make numerical assumptions starting from such a low sample. However, the information collected

allows to make some interesting qualitative observations, presented in the following lines, together with the most significant cases in this perspective.

The value, when present, generally ranges from hundreds to a few thousand users involved. It is possible to observe that the adoption is very low, considering the potential, once again indicating the embryonic state of the practical application of the SSI model. It is clear that research on how to make it scalable is still ongoing.

Among the cases in which the diffusion of the model is known, three particular projects are presented. The first two have some similarities that allow to show some common aspects and limitations. These two projects are both developed in Switzerland, the first is called Zug ID and arises from the collaboration between the city of Zug and uPort, while the second, called Schaffhauser eID+, arises from the collaboration between the canton of Schaffhausen and Procivics. Both are born as eGov projects, which are solutions that involve the use of digital identity as an alternative to the traditional system to access some public administration services. Both are motivated by the attempt to restore control of their own identity to the citizens, also guaranteeing high levels of privacy and security, enough to allow access to government services, which generally require a high LoA. In terms of technology and user experience, both projects involve the use of a specific mobile app with at least two authentication factors required. The onboarding is carried out in the presence, directly in the local government offices. Personal documents verified in person are the data type associated with the identity profile. The main difference between the two cases and also one of the main reasons for the interruption of Zug ID, lies in the number of services accessible with the solution offered in the two cases. In the case of Zug ID there were only two accessible services, elections and bike rental, while the eID+ case boasts more than forty different services accessible with this identity. This has allowed a much more numerous adoptions, over two thousand citizens (about 2.4% of the population of the canton), still increasing as the project is operational, compared to the 267 citizens who have adopted Zug ID. This highlights once again the importance of expanding the operability of SSI to as many application areas as possible, so that it is not one of the many other identities to be managed by the end user, but instead is a unique alternative to the traditional models.

The last case, on the other hand, is presented because it is very representative of how many of the projects surveyed have been developed. This is Xride, a project carried out in Germany, which sees Jolocom, T-Labs and other partners among the participants. The idea behind Xride is to decentralize the system that allows scooter sharing, relying on the SSI technologies provided by Jolocom. In the pilot functionalities such as identity, payments, and charging, were fully decentralized, unlike what happens in the traditional systems. This allowed to have a less costly, more secure, and more efficient scooter sharing system, with benefits for both users and providers. The interesting aspect for the dimension of analysis considered is that the project involved a group of Deutsche Telekom employees, who played the role of testers, providing advice and observations on the functioning of

the proposed system. The exact number of employees involved is not known, however it is a classic example of how different projects have developed. They started with a partial solution, tested by the employees of the companies involved, in order to further develop the project, before landing on the market. In the case of Xride, the T-Labs operating unit has hundreds of employees, even if the actual users are only a part of them. So, in this case, there are numbers lower than the two previous cases, which is to be expected since the solution is used in a limited environment. This phase can be very important, as it allows developers to find problems that did not emerge during the theoretical design and at the same time verify that the solution is actually valid for users. Obviously, this does not guarantee success on a commercial level, but it certainly helps to be more oriented to the needs of users.

3.1.5 Economic Sustainability

The data on the economic sustainability of the SSI model are often only partial and are present for a low percentage (33% of the total) of cases too. Despite this, it is still possible to make interesting qualitative observations both on a general level and by going into detail with some cases considered significant.

The most immediate observation that can be made, following the results of the analysis, is that, as it is still an uncertain and developing system, there are no established business models. And even in the few known cases, the economic model has been adapted as some cases evolved, since economic sustainability depends on the actors participating in the case and the number of users. In addition to this, it should also be noted that in most cases the SSI system relies on other technologies, such as Blockchain and Wallets, which are also still being defined in terms of business model.

Looking at the composition of the ecosystem and the main actors involved, already described in Chapter 1, it is possible to identify different business models that involve them. These models generally involve two of the three actors, or all three. For example, in many traditional cases the user pays the issuer to receive a certain verifiable credential. Or the issuer and the verifier pay fees to the platform hosting the project. In other cases, the exchange of economic value takes place between the holder and the verifier. Charging any party represents a real risk of lowering the adoption of the model. However, it is clear that, in order to survive, the system must also be economically sustainable. As the principles of the SSI are defined and to effectively mark a departure from the traditional models, the holder should not be the paying party. Furthermore, this favors adoption by people and consequently also attracts the other actors present in the ecosystem. Precisely this aspect emerges from the analysis of the cases in the census. Indeed, the actual business model adopted is often not specified, but it is openly stated that the user will not have to pay any amount.

In addition to what has been mentioned so far, it should be remembered that a further difficulty in the economic sustainability of the SSI model lies in the fact that as it is designed it is not possible to sell user data. In traditional models, the data belongs to the platform or to issuers and verifiers, who can treat it as they want. In the SSI model, data is held only by people, any action that concerns them requires their consent. Another consequence of this peculiarity is that it is not even possible to carry out targeted advertising campaigns since the platforms do not have user identity profiles. Furthermore, again due to the need to receive consent from the holders, in the SSI model it becomes even more difficult to do matching or two-sided markets. Currently these are the most popular and profitable business models in the industry, their impracticability makes it more difficult to find a profitable business model for the SSI system. On the other hand, it should also be emphasized that this once again certifies the great potential of the model in improving the protection of the interests of the end user.

More in detail, three significant examples are presented. The first case presented is DIZME, which is a decentralized identity network, born in Italy and currently active. Thanks to DIZME it is possible to build an identity profile with a high LoA, which can be used to access the services offered by the participating SPs. In terms of process, it is very similar to the majority of cases, relying on blockchain technology, on a mobile digital wallet. Regarding identity attributes, it is possible to attribute certified personal documents to the identity profile, together with self-declared contact and biometrics information. The peculiarity of this case lies precisely in the economic model chosen. In the DIZME case, it is the verifier who pays the issuer, this makes the system free for users, and economically recognizes the value of the credentials issuer. Indeed, the issuers, which will make available their credentials to the holders undertaking the liability, the identification and classification costs, will be reimbursed by the verifier, who will enjoy several benefits introduced by this system. These include a faster, more secure, and cost-effective credentials verification mechanism. The economic model proposed by DIZME is quite interesting, as it is theoretically aligned with the concept of SSI and rather advantageous for all the parties involved. However, it should be noted that the functioning of the rewards system exploited by the case raises some doubts about the effective protection of the user's privacy. Because the rewards are paid directly by the verifier to the issuer and to the end user, as the latter initially anticipates the Transactions and Sign Request fees. This could lead to a violation of the Protection principle of the SSI model.

The second case is MemberPass, a project active in the United States, although at the moment limited to members of the Credit Union only. MemberPass provides a way for credit unions to quickly and confidently verify their members' identity, providing an improved member experience, reducing fraud, and increasing operational efficiency. In terms of economic sustainability, even MemberPass does not require any fee on the member, who in this case has the role of end user. However, it requires payment from the credit union that wants to use its technology. The payment does not have a fixed fee but is customized from case to case. In this way MemberPass recovers its

costs without weighing on the users, but on the SPs who take advantage of the benefits offered by authentication through this system. Similarly, entities that rely on the technologies offered by the Sovrin and uPort protocols to develop their projects, must pay a certain fee to be able to use these technologies on an ongoing basis. The difference is that in this case there is a well-defined tariff plan that depends on the type of project. For example, Sovrin proposes lower prices for projects under development, compared to those of active projects. However, there are exceptions such as Identity For Good, a project developed by the Red Cross in collaboration with Evernym, World Economic Forum and Sovrin. The aim of the project is to give a permanent and directly controlled identity to the people most in need, in order to facilitate aid distribution operations. In this case, given the humanitarian purpose of the project, no payment was requested, neither for development nor for maintenance during the test phase.

What seen so far basically confirms the fact that it is still too early to understand what the winning economic model is. Furthermore, looking at the numbers it is evident that this is not one of the priorities and is a factor ignored not only by literature, but in some cases also on a practical level. However, there are also cases in which the economic sustainability of the model was considered, adopting specific business models. It is difficult to predict with so little data how the SSI model will evolve as it matures, however a trend that emerges quite vigorously from these results is that most likely, at least in the initial phase, the payments will not be on the end user.

3.2 SSI PRINCIPLES

3.2.1 SSI Protocols

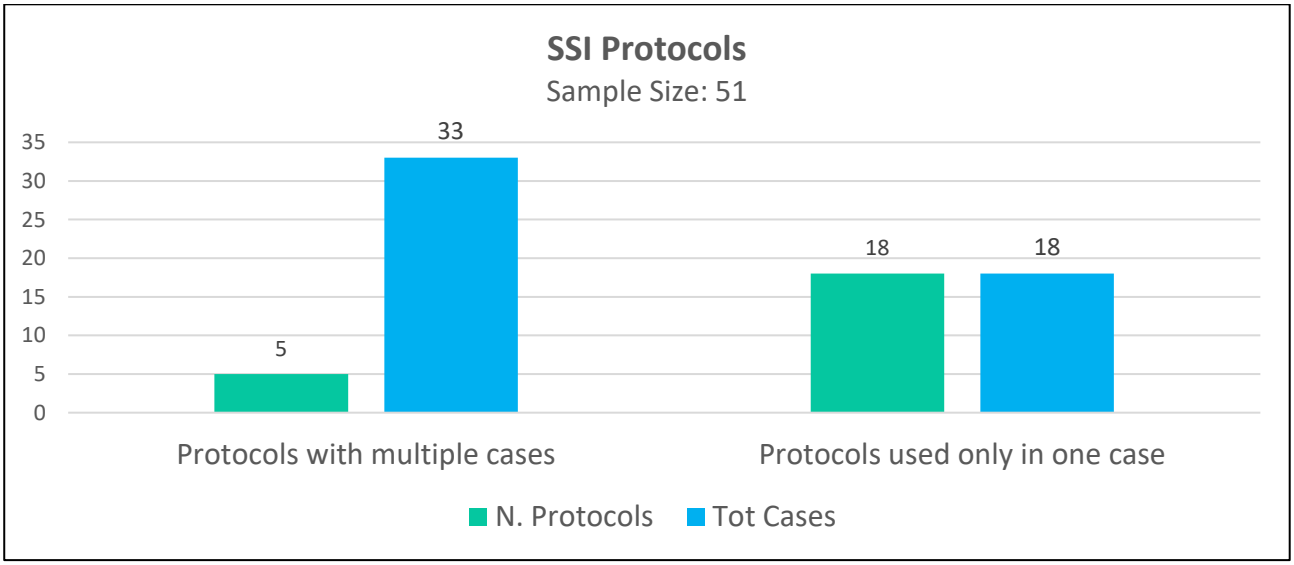

Figure 3. 6 SSI protocols

It is possible to observe in *Figure 3.6* that there is a multitude of protocols, as many as 23 for 51 cases. Of these, only 5 (Sovrin, Jolocom, W3C Standards, uPort and STONledger) are used for multiple use cases, but they cover 65% of the total cases. This is an interesting fact, which tells two things: the lack of a precise international direction implies that most projects try to develop their own protocol, especially when there is a geographical and cultural distance from the more established ones. On the other hand, there are also some protocols that are making their way and establishing themselves as the main ones in the landscape.

Considering the projects that involve the institutions, there are no significant differences in the trend. In fact, 70% of cases involving institutions rely on protocols with multiple cases of application, compared to 65% overall. Rather, it is interesting to note that uPort, which is a multi-application protocol, has only been involved in projects that included government agencies. Two of these cases are for the development of an SSI usable in the eGov field, while in the other case for the verification of the identity and role of workers on large work sites of public companies.

The three main SSI protocols that emerge are Sovrin, Jolocom and W3C Standards. They alone account for 55% of total cases. Each of them has peculiarities, which in a certain sense motivate their success, and make them particularly attractive and interesting for the future. For this they will be described in detail in the next lines.

The **Sovrin** protocol is the one with the largest number of cases. It is a public service utility enabling SSI. The Sovrin Network is decentralized, meaning individuals can collect, hold, and manage identity credentials, without relying on individual centralized databases that manage the access to those credentials. Sovrin is an independent organization that is responsible for ensuring that the Sovrin identity system is public and globally accessible. On a technological level, Sovrin is based on a very particular blockchain, as it is public, but permissioned. It also implements Privacy by Design on a global scale and selective disclosure of personal data using Zero-Knowledge Proof cryptography. Sovrin can be considered to all intents and purposes a supplier of a network in which to operate within the SSI. This network boasts several participants. In terms of stewards, namely trusted organizations, which are responsible for operating the nodes that maintain the Sovrin distributed ledger. But also, in terms of partners with a different role. Another interesting aspect is that Sovrin can be used in many different areas. Indeed, the projects that use it as protocols are very varied and have allowed the organization to build a wide experience and identify different insights on the practical side of the SSI model.

Jolocom is an open-source protocol for people and organizations to create and interact with SSI. The Jolocom Protocol facilitates the generation and management of DIDs, VCs, and cryptographic signatures, which are the core building blocks of Jolocom identities. Jolocom identities are created entirely locally using hierarchical deterministic keys and are designed to enable management of

multiple personas by individual users as well as preservation of pairwise anonymity in context-specific interactions. On a technological level, Jolocom relies on the Ethereum blockchain, which is public and permissionless and has developed its own wallet, in the form of a mobile app. Despite this, it still leaves the possibility for users to select a different wallet. Following the analysis of the protocols, it can be said that Jolocom is certainly the most aligned and attentive to being compliant with the SSI model. In the documentation found during the research, many technical choices regarding the Jolocom cases are justified using the ten principles of the SSI. Unlike Sovrin, Jolocom, in addition to playing the role of supplier, often has a more supportive role in the projects in which it is involved, participating directly in the development phase, and helping to shape the final solution. It can be said that the cases using Sovrin and W3C Standards, a protocol that will be described later, the projects are often stand-alone and require specific process technologies. Instead, in cases using Jolocom, the projects are still separate, however the process technology is the same and the information can be stored in the Jolocom wallet, even if the use of other wallets is possible. So, it can be said that Jolocom is trying to create an integrated ecosystem, among all its use cases, while leaving the user the freedom to eventually use alternatives. However, it should be emphasized that at the moment there are no active cases using the protocol. Although many projects in the experimental phase involve a certain number of users, acting as testers, selected from a limited context, such as employees of the companies involved in the project or some volunteer citizens of the participating cities.

Finally, the **W3C Standards** protocol is presented. In this case, unlike the previous ones, it is not a real organization that directly provides a protocol. Rather, the World Wide Web Consortium (W3C) is an international non-governmental organization which aims to promote the development of all the potential of the World Wide Web and spread the culture of accessibility of the Net. In order to succeed in its intent, the main activity carried out by the W3C is to establish technical standards for Web technologies. Among these is also included the SSI, the cases developed according to this protocol will therefore follow the standards and practices defined by the W3C, such as those for DIDs or VCs. It is clear that as anticipated the cases that refer to these directives are completely independent from each other and do not necessarily receive direct support from the protocol provider. In the face of these disadvantages, which among others disfavor the passage of the experience accumulated in other practical cases, there are also some advantages. Indeed, by itself, standardization ensures that the web works equally well for everyone, regardless of their location or technology. Furthermore, W3C standards also improve issues of accessibility, privacy, security, and internationalization. All aspects, which favor compliance with the SSI principles, as well as its dissemination since W3C standards are available free of charge. The lack of standards defined for the SSI model was one of the problems identified by the literature, certainly the work of the W3C tries to solve this problem, facilitating the growth of the SSI system.

Precisely for these three protocols a cross analysis was carried out with the application areas of the cases that use them. These results are anticipated and shown in *Figure 3.7*, while the general results on the application areas in the overall case will be presented later in the dedicated section.

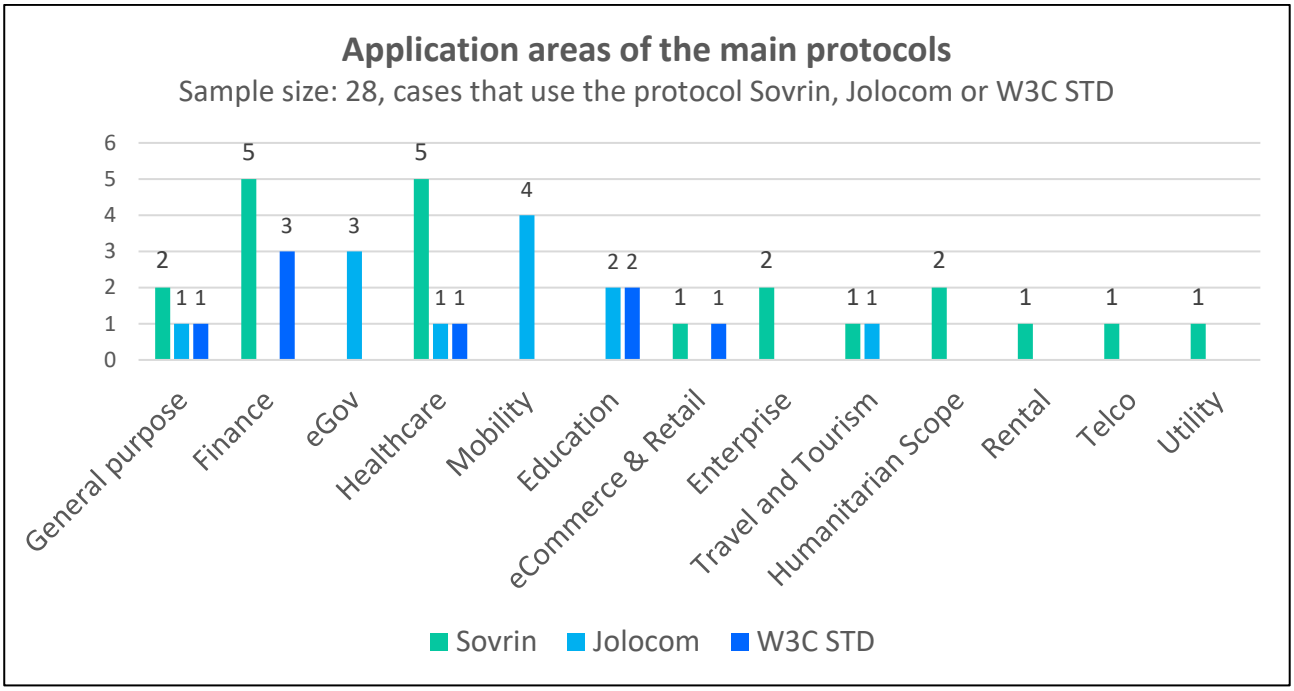

Figure 3. 7 Application areas of the main protocols (each case can have more than one application area).

It is possible to note that all three main protocols have at least one case where they are meant for general application. Furthermore, each of them is used in at least 4 different application areas. This confirms their ability to cover different application areas and not be limited to specific niches or sectors. This is very important to ensure compliance with the interoperability property and more generally to offer an identity system that can be used in all areas. This could allow users to have a single digital identity valid in every context. Looking in detail at the sectors, only healthcare has at least one case for each of the three protocols. As for the other sectors, Jolocom is more present in projects involving government bodies and on mobility. While Sovrin and the W3C Standards are mostly used in finance and business.

3.2.2 SSI Principles

The analysis for each of the ten SSI principles in practical applications led to some interesting outcomes. First, it should be noted that the information relating to this dimension of analysis was not always present or lacked the level of detail necessary to be able to verify all ten principles. However, in 88% of cases the necessary information was available. An interesting aspect that

emerges from a general analysis is that there are properties that are more easily respected than others. In fact, all cases comply with the principles of **Existence**, **Access**, and **Transparency**. It must be said that these are three fundamental principles, but rather simple to respect as they are defined. Furthermore, only one case shows criticalities in compliance with the **Control** and **Consent** principles. This is the iRespond project developed in Thailand to give a Self-Sovereign Identity to the refugees present in the Mae La refugee camp. Through their digital identities, participants are able not only to access improved healthcare services but also securely store educational and professional credentials. During the development of the project these two principles were stated, however given the complicated context they may not always be effectively respected. In any case it is an excellent result, indeed, Control and Consent are two peculiar principles of the SSI theoretical model, the fact that they are respected in the vast majority of cases indicates that even on a practical level there is great attention to these aspects. Similar speech for the principle of **Protection**, it is respected in all cases except two, due to the selected economic model which could cause the infringement of some rights of the user.

The remaining four principles, on the other hand, are those in which the greatest criticalities have been discovered. Regarding the principle of **Persistence**, it must be considered that despite the incentives for the adoption or search for a sustainable business model, there is no guarantee that each solution will be used and / or will exist forever. While unlikely, it is not impossible for a network to fail. However, in this case it is necessary to make a distinction: the more adopted and representative cases of the ecosystem, the more difficult they will risk failing. For example, Jolocom is based on the Ethereum platform which can be considered stable, given the high economic value incorporated. Likewise, Sovrin includes, among its partners, large companies, such as Deutsche Telekom, InfoCert and IBM, which have every interest to keep the ecosystem alive. Instead, in cases where the ecosystem is limited and managed by a small consortium or by individual companies, the guarantees on the future persistence are much lower. To conclude, it can be said that generally the principle of Persistence is guaranteed especially in cases that rely on more established protocols, indeed, not only they are more stable ecosystems, but often also leave the control of the DID to the user himself. The fact that DID continues to exist regardless of what happens to the ecosystem implies the persistence of the identity itself. It follows the principle of **Minimization**. The analysis of this principle shows interesting results. Indeed, in many practical cases (93%) it is considered one of the most important principles and compliance with it is clearly stated and illustrated. However, in 7% of cases this principle is not respected, and the user is forced to disclose more personal information than necessary. Of these two cases are active. The principle of Minimization is fundamental for the full achievement of the potential of the SSI model and should be taken into great consideration during the development of the project. The fact that active projects do not respect it is a wake-up call, indeed, it shows that the principle itself is not necessary for the functioning of the system. For this reason, more attention is needed in complying with it, only in this way all the benefits promoted by the SSI model can be guaranteed to the users. Finally, there

are the two most critical principles, namely **Portability** and **Interoperability**. In the case of the **Portability** principle, the major problems that arise in analyzing the results concern the limitation to one or a few alternatives for the users. More specifically, identity profiles are often only linked to the only wallet supported and / or developed in the project. This explains why the principle of Portability in 67% of the cases is not respected or is only partially respected. In the first case, only one alternative is possible, generally the wallet or mobile application adopted in the project. In the other cases more wallets are available, but the choice is limited to those indicated by the developers. This is a problem often recognized by the creators themselves, but currently unsolvable due to technical or commercial limitations. There are cases in which this problem has been solved by indicating a reference wallet, but leaving the user free to choose, as happened in some cases related to the Jolocom protocol. Furthermore, established protocols are also working on a way to make identities operable within each of their ecosystems, in order to favor portability. In other cases, developers are working or will work to expand the number of compatible wallets. The most critical cases are those in which a specific application has been developed, perhaps due to some peculiarities of the project, without considering portability. Portability is a fundamental property in order to guarantee the longevity of the identity and avoid lock-in effects. Despite the critical issues, the recognition of the problem and the attempts to solve it are encouraging and bode well for its resolution, especially with the maturation of the SSI model. Finally, as regards **Interoperability**, the main limitations derive from the fact that often the cases considered are still immature, with a limited ecosystem both in terms of participants and in geographical terms. This makes it more difficult to assess actual compliance with the principle. More in detail, in 78% of cases, interoperability, as defined in the theoretical part, is not respected or is only partially respected. However, it should be specified that in many cases this is a recognized problem, and potentially solvable by expanding the ecosystem both in terms of participants and application areas, and in geographical terms, making the identity usable even outside the environment in which it has been developed. Obviously, this is the growth objective of many projects, even if its actual achievement is not taken for granted. A check on compliance with this principle can be carried out in the future when the cases will be more mature.

The difficulties encountered in several cases in fully respecting all ten principles confirm that the SSI model is still predominantly a theoretical model. In addition, a minority of cases even ignore some of these principles, indicating an interest more dictated by the hype of the moment, than by a real interest in the ideas behind the SSI. However, it should be emphasized the presence of cases and protocols, including those three presented in detail in the previous paragraph, capable of respecting all ten SSI principles and making technical choices motivated precisely by adapting to the theoretical dictates of the model. This confirms that a transition is taking place towards the practical application of the SSI concept. Furthermore, many of the problems encountered in complying with the most critical principles could be overcome once the model reaches a certain stability and maturity. To do

this, multiple attempts are needed, but above all investments to attract people's attention, providing them with a solution that effectively brings the theorized benefits.

3.3 TECHNOLOGICAL ASPECTS

Thanks to the analysis on the technological aspects of SSI cases, it was possible to identify the main trends and the most established practices in the technological context of the SSI. Although not always present, the information found has allowed to take a rather precise picture of the technological perspective, given the presence of several predominant aspects.

3.3.1 Blockchain

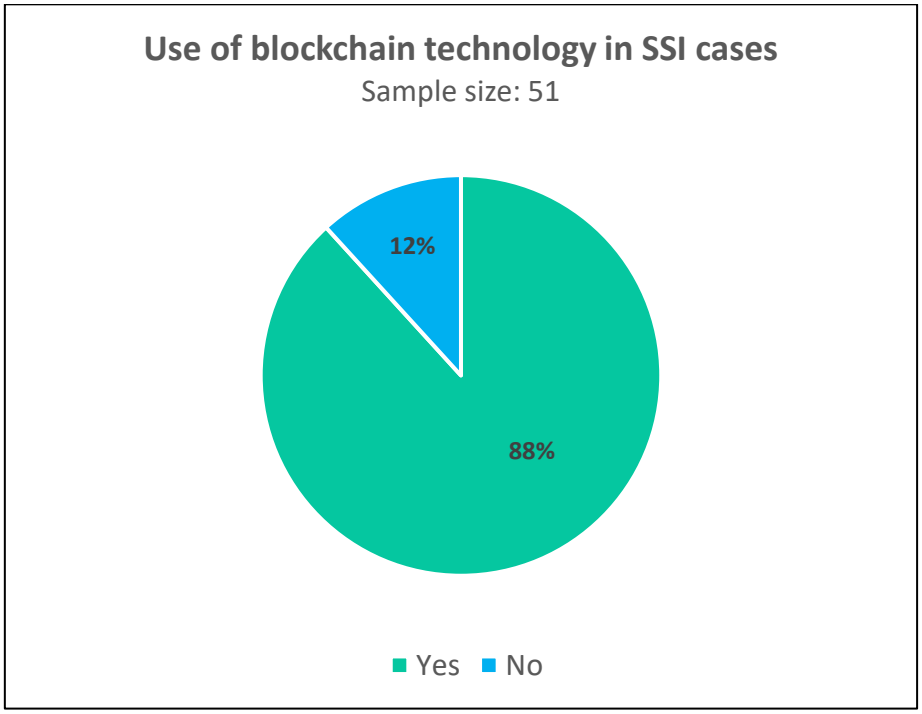

Figure 3. 8 Use of blockchain technology in SSI cases

As mentioned several times, blockchain technology is not essential for the functioning of the SSI system, however several advantages arise from the combination of these two entities. A strong relationship between the two concepts has already emerged from the analysis of the literature. This link is also confirmed in practice, as shown in *Figure 3.8*, indeed, 88% of cases use this technology.

As for the platforms, there are several in the various cases, however the most used are Ethereum and Hyperledger Indy, which cover about half of the cases. There are also a few projects where the blockchain has not yet been selected, but the characteristics it should have are specified.

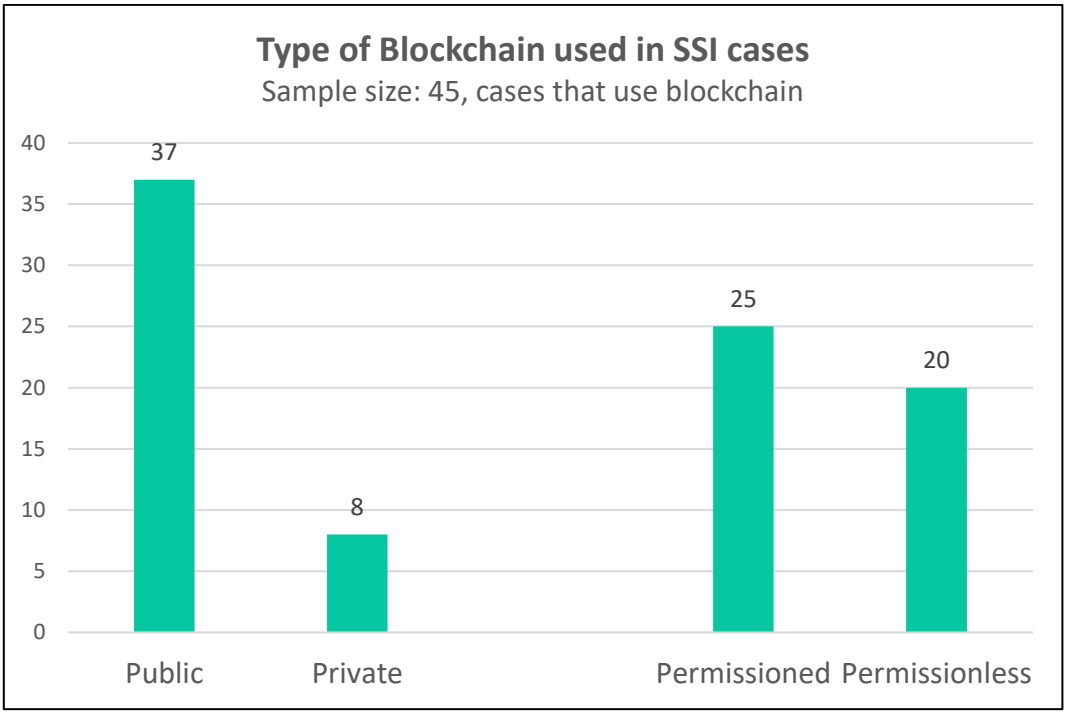

Figure 3. 9 Type of Blockchain used in SSI cases

As regards the characteristics of the blockchain used, whose data are represented in *Figure 3.9*, there is a clear prevalence of public blockchains, however this does not correspond to a prevalence of permissionless blockchains, which are, instead, the minority. This result is influenced by the numerous cases on Sovrin that rely on a public and permissioned blockchain. This is an interesting result that allows to make two observations. The ideal case for the SSI system is the use of a public and permissionless blockchain. Only in this way, it is possible to avoid a return to a sort of centralized model, in which some rights are owned by only one part of the people in the ecosystem. A public and permissionless blockchain as described in the dedicated chapter, allows anyone to operate on the platform without having to go through a controlling authority. The results show that the prevailing trend is that of granting access to everyone, since the blockchains used are mostly public. However, this does not also correspond to the freedom from authorities that tend to centralize and control the blockchain, which in fact are mainly permissioned. This represents an obstacle to the full achievement of the potential of the SSI model. And it is also one of the main advantages of the Jolocom protocol compared to Sovrin. The first, indeed, uses Ethereum, one of the most established blockchains, which is public and permissionless. Sovrin, instead, as already mentioned, is based on Hyperledger Indy, which is public but permissioned. However, it is necessary to make a clarification, the presence of this conglomerate of power does not necessarily lead to a centralized model, indeed, Sovrin itself for example guarantees that it is more of a guarantor than a controller role,

however the risk is potentially present. From this point of view, the results of the analysis carried out on the 10 active cases that use blockchain technology are to be considered much better and encouraging, as shown in *Figure 3.10*.

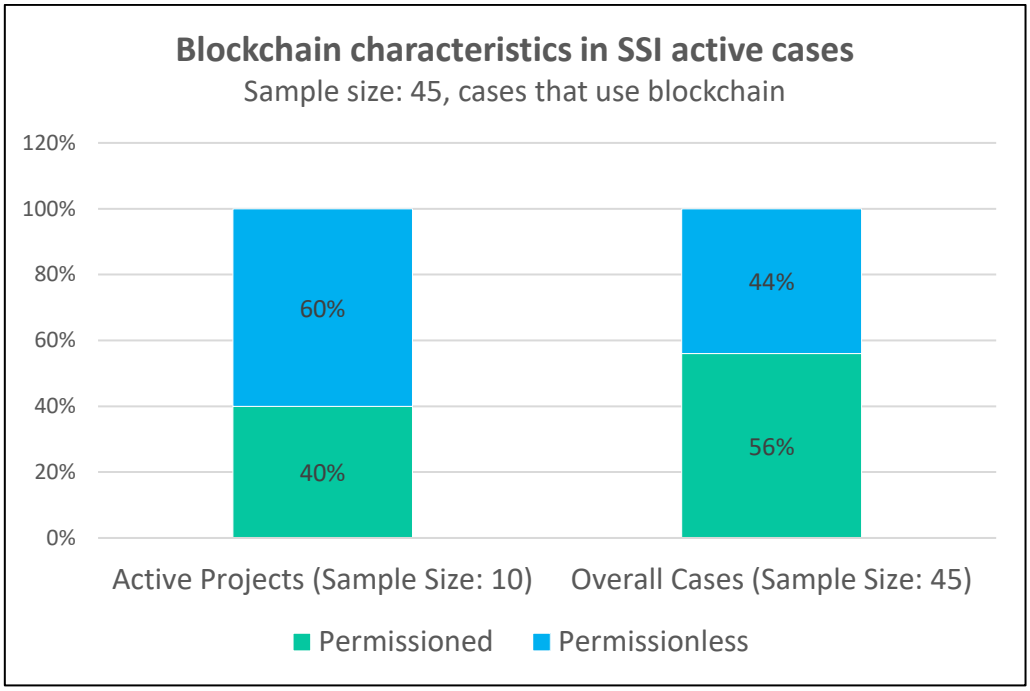

Figure 3. 10 Blockchain characteristics in SSI active cases

Indeed, there is a prevalence of permissionless cases, compared to permissioned ones, the opposite of the general results. This is an interesting result, given that the permissionless feature of the blockchain, like already said, is the most suitable for the SSI model, the fact that it is prevalent in active cases is a symptom of its effective applicability.

An analysis was also conducted on the 6 cases that do not use the blockchain. Among these cases it is possible to make a distinction between those that consider the use of the blockchain, even if it is not implemented right now. And those who, on the other hand, have already thought of a solution without it.

In the Schaffhausen eID + case, already mentioned, the solution was never based on the blockchain. However, some tests were conducted[14] to verify its applicability in context. It can therefore be said that the use of the blockchain has been considered and is an open option for the future. Similarly, in the Tangem case that uses NFC cards, there is no single alternative on the technology to be used for the verifiable data registry. Among the options the blockchain is mentioned, but it is also possible to use other decentralized networks, unspecified. So also, in this case the use of the blockchain has been considered, but it is not the definitive solution for the moment. However, the functioning of the system is similar to the blockchain case. Also in the case of Mattr VII, a project developed in New

[14] https://sh.ch/CMS/Webseite/Kanton-Schaffhausen/Beh-rde/Services/Schaffhauser-eID--2077281-DE.html

Zealand and still active, the blockchain is among the possible options. More specifically, DIDs employed in this system can be stored on a variety of different data registries, such as blockchains and public databases. A limitation of this case is that it is not clear who the entities collaborating with Mattr are. For this reason, it is not clear what the choices of the implementers of this project are, in terms of data register. In any case, the functioning of the system is similar to that of the systems that use the blockchain, which, indeed, appears among the options.

An interesting case, to understand how these blockchain free projects work, is IRMA. In this case, the credentials are cryptographically linked to a mobile phone and to each other via a secret cryptographic key. This personal private key is critical to the security of the IRMA app. Therefore, having only local storage is not enough to guarantee the necessary security. This because the phone can be rooted or hacked. As a result, the developers thought about storing a portion of the private key outside the phone on a so-called keyshare-server. This server is managed directly by the organization behind IRMA. To reveal the attributes, the user must give consent and enter the PIN. The IRMA PIN code is checked by the keyshare-server. If the PIN is correct, the server will participate with its own portion of the personal secret key and the attributes can be disclosed. The keyshare server in any case will not see the attributes themselves, nor to whom they are disclosed. So, in this case the role of the blockchain is played by the IRMA server. It verifies the attributes and returns their status to the requestor, after receiving the consent from the user. Basically, on the IRMA server there is a part of the secret keys and the information about the attributes necessary to carry out the verification.

In the remaining two cases the use of the blockchain is not mentioned, but there is no information available on how the system works. More generally, by analyzing the cases without blockchain, no particular differences emerge from the cases that use this technology. The results crossed with the various perspectives are similar to the overall case. Including results on the analysis of the ten SSI principles for these cases. There are, indeed, some issues in the principles already identified as critical, such as portability and interoperability.

3.3.2 Integration Technologies

The integration technologies are generally a set of specific procedures or tools designed to solve an integration problem between different software or between different software components. Similarly, this category includes the authentication protocols that are integrated into the solutions developed to be able to verify the identity of the users who use it.

The main protocols, whose use in practical applications has been verified, are SAML, OIDC, API & SDK. For the description of these concepts, the reader is referred to the literature chapter. The

results show a clear trend, namely the predominance of API & SDK, in the few cases where other technologies are used it is only to allow retrofitting with existing systems. An example of this is the Tango project, which is being developed by the city of Buenos Aires with the collaboration of some private entities. In Tango's Whitepaper, it is stated that for backwards compatibility with Web 2.0 applications that use Federated Identity mechanisms, specifically for SSI Login scenarios, the use of the OIDC protocol could be considered.

The use of the APIs & SDKs guarantees greater flexibility, and they can be easily downloaded, modified, and updated. Furthermore, they permit to create different experiences for users, letting protocols, functions, and commands be adapted according to specific demands. This favors customization. However, the biggest advantage is the ease and variety of applications they can be integrated into, facilitating the dissemination of the SSI model.

3.3.3 Process Technologies

Information on process technologies is known and well defined in 75% of cases. To these must be added 12% of cases in which the available data are not complete. In the few remaining cases the information is not available or has not yet been defined within the project. The analysis on process technologies shows that in 79% of cases where information is available a wallet is used, while a mobile application is present in all cases.

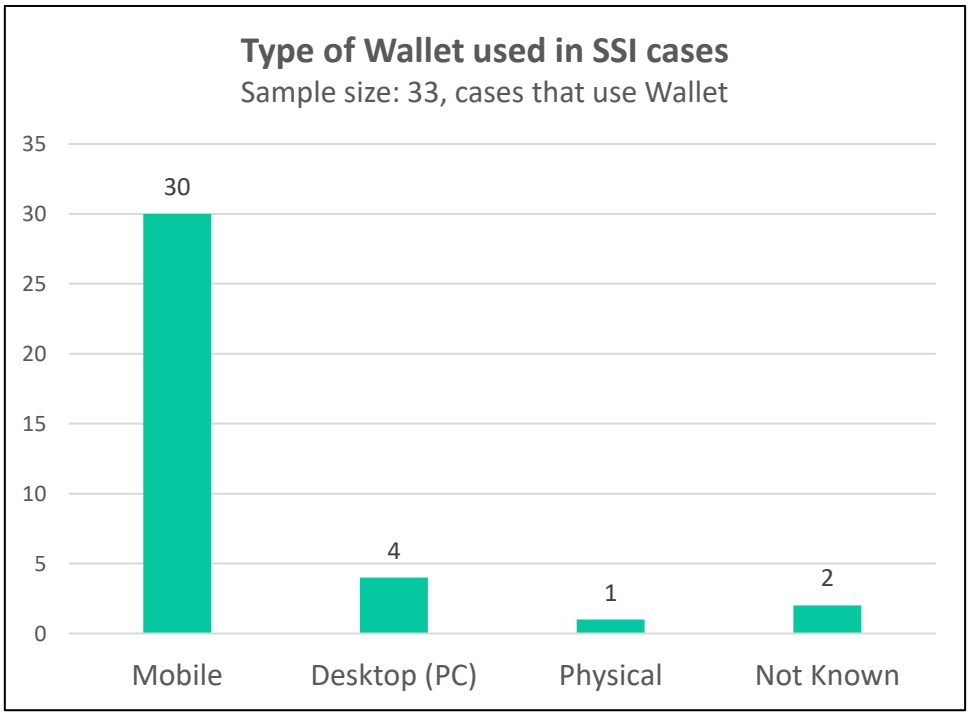

Figure 3. 11 Type of Wallet used in SSI cases (the supported wallet can have multiple versions)

More in detail, *Figure 3.11* shows that in 97% of cases mobile wallets are used. Desktop wallets are never used alone, but always as an alternative to a mobile one. The physical wallet is used in a single case, exploiting NFC Cards. There are two major benefits to using mobile wallets and mobile applications. The first is portability and the second, most interesting for the SSI model, is the ease with which they can be obtained. Indeed, nowadays many people have a smartphone on which to install them. This can greatly facilitate the access and diffusion of the SSI system.

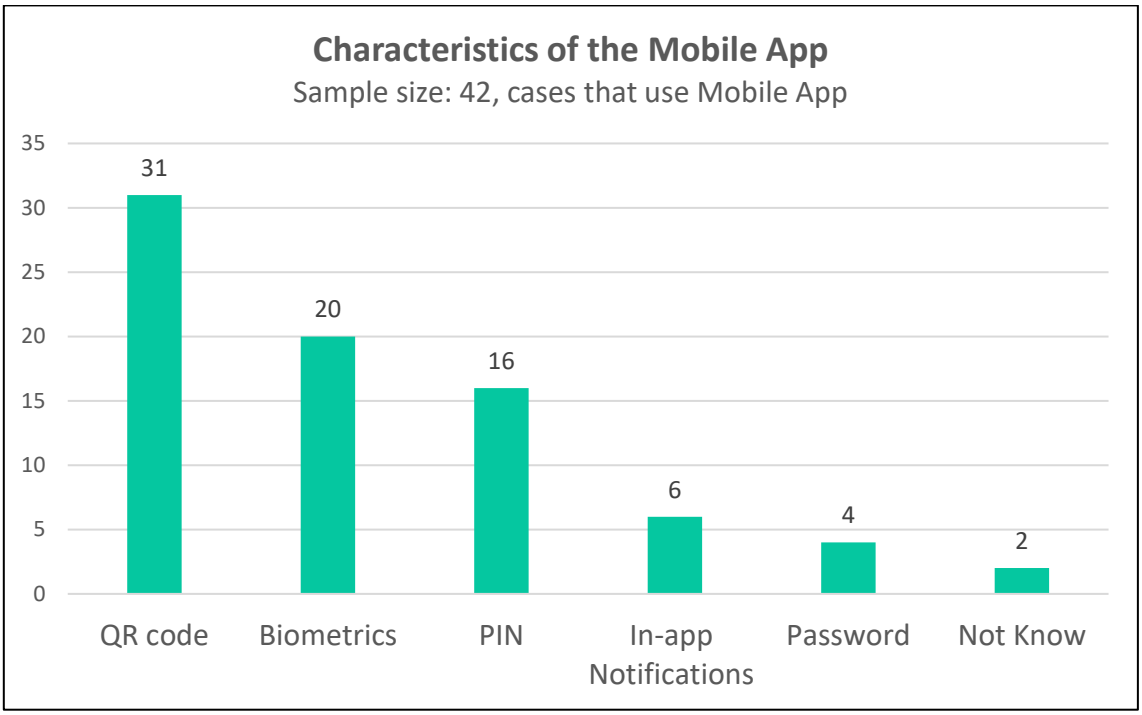

Figure 3. 12 Characteristics of the mobile app used in SSI cases (each mobile application can have more than one of these options)

As for mobile applications, the results in *Figure 3.12* show that the use of QR Code is predominant. Methods such as PIN and Password are still quite widespread, but in sharp decline if compared to the traditional systems, especially Password usage. Biometrics is widely used and supported. However, its use is optional in 40% of cases. In addition to the observations in technological terms, these results allow us to make others relating to the user experience, described in the next section.

3.4 USER PERSPECTIVE

3.4.1 Access and use credentials

The results on the factors required for authentication are presented in *Figure 3.13*. Access and use credentials information is known in 67% of cases. In one case they have not yet been defined. While in the remaining cases the information is not available.

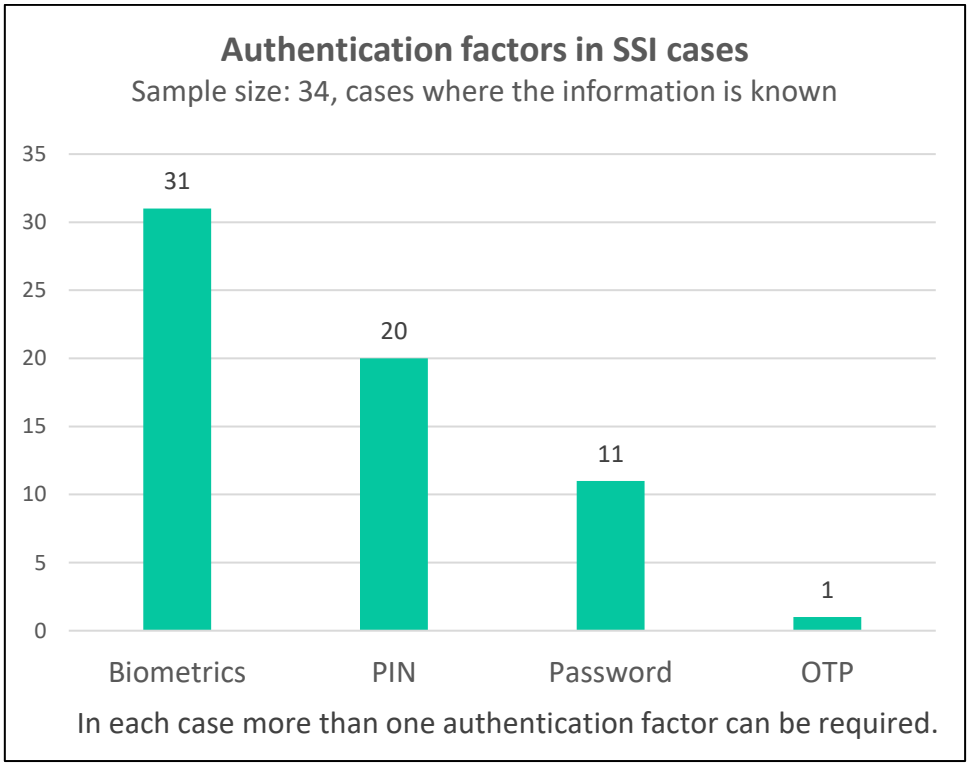

Figure 3. 13 Authentication factors in SSI cases

The access methods return some interesting data. First, there is a decline in the use of User ID and Password. This trend is similar to that identified for process technologies. Secondly, the most used methods are biometrics and PIN. However, in the case of biometrics it should be highlighted that it is not to be used compulsorily but is supported as an optional option for extra security in 48% of cases. Overall, these findings mark a major shift in the user experience. Indeed, SSI projects mark a shift from the methods used in traditional systems, such as username and password, towards more innovative technologies. These technologies are safer, but still very simple to use for the end users, improving their experience. To further increase safety, MFA is introduced in some cases.

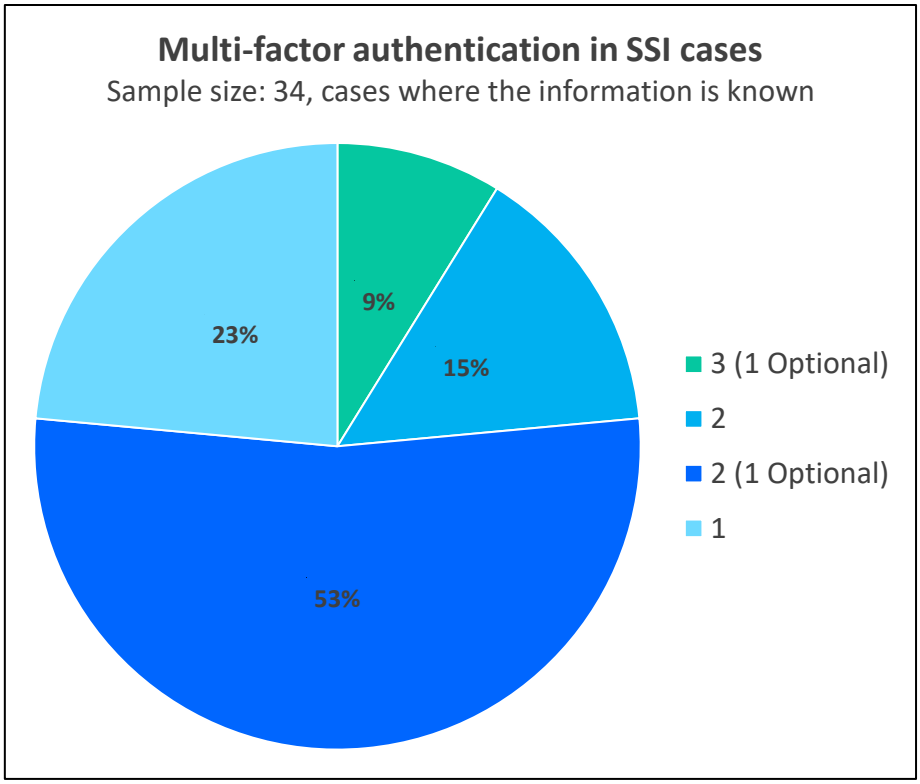

Figure 3.14 Multi-factor authentication in SSI cases

As shown in *Figure 3.14*, despite its importance, MFA is compulsorily used only in 24% of cases. Number similar to that of cases where just one authentication factor is needed. When only one authentication factor is required, in 75% of cases it is biometrics. The remaining cases are equally divided between PIN and Password. However, the fact that MFA is the direction of the future is evidenced by the fact that 77% of cases support it. The percentages remain similar even in cases involving institutions.

3.4.2 Onboarding

The information on the Onboarding procedures is known for 69% of cases. In one case they have not yet been precisely defined. While information is not available for the remaining number of cases.

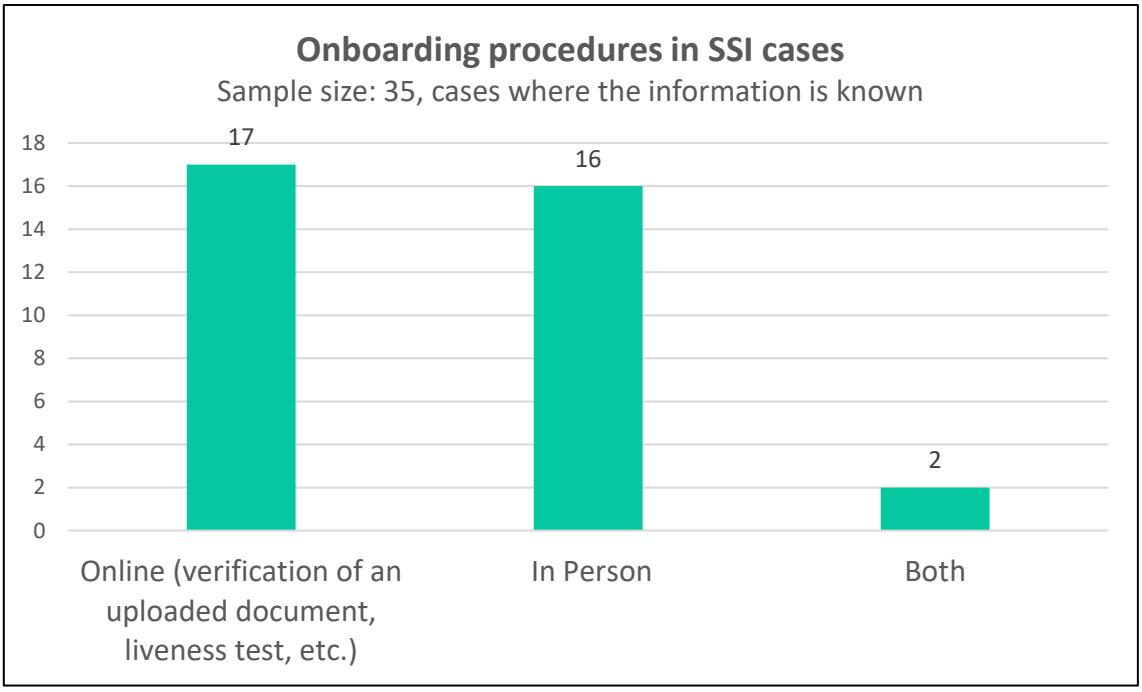

Figure 3. 15 Onboarding procedures in SSI cases

Figure 3.15 shows that in person recognition of people remains a very popular method and is often required for a higher LoA. However, online onboarding procedures are growing and are present in greater numbers than those in presence. From a user experience point of view, in person verification is similar to what currently happens when a person requests a new document or needs to renew an expired one. However, online onboarding would reduce the time and costs required for the procedure. In addition, it would facilitate access to identity for people residing in rural areas or with difficulty moving around. This would improve the final user experience, as well as certify an additional benefit deriving from the adoption of the new model. However, considering the cases in which institutions are involved, the results change drastically. Specifically, in 80% of cases, where the information is known, in person onboarding is required. Additionally, such projects account for 75% of the total cases where in person onboarding is required. The result is not surprising considering that the services offered by these entities require a high LoA. However, it should be emphasized that this could be a limitation to the potential of SSI. Indeed, generally the documents issued by the institutions are fundamental for building reliable identity profiles. The results show that to obtain them it is still necessary, in most cases, to carry out in person recognition.

There are different types of online onboarding. The most common is the verification of documents uploaded by the user, which is facilitated by the use of the blockchain. There are also other methods of online onboarding, often associated with specific cases and used in combination with the other methods. Among these, there are the liveness test, the verification of mobile number or of email.

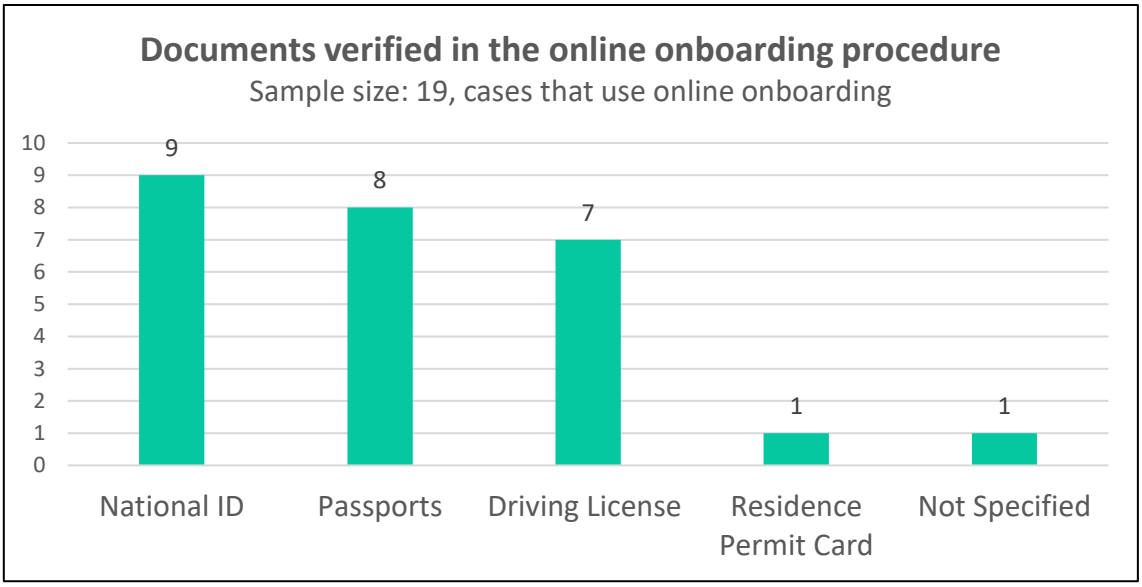

Figure 3. 16 Documents verified in the online onboarding procedure (multiple documents can be verified per case)

Figure 3.16 shows that the documents checked most often are the most common ones, such as National ID, Passport and Driver's License. This should not cause any problems for the user since these are the most easily accessible documents. At the same time, this confirms what was previously stated. In other words, to build a reliable identity profile through online onboarding, documents issued by the institutions are required. However, these documents are mainly issued following in person recognition, with the consequent limitations described above.

3.5 DATA

The type of data relating to the SSI cases surveyed are known in 94% of cases, although in 2 cases some alternatives are proposed, but not definitively.

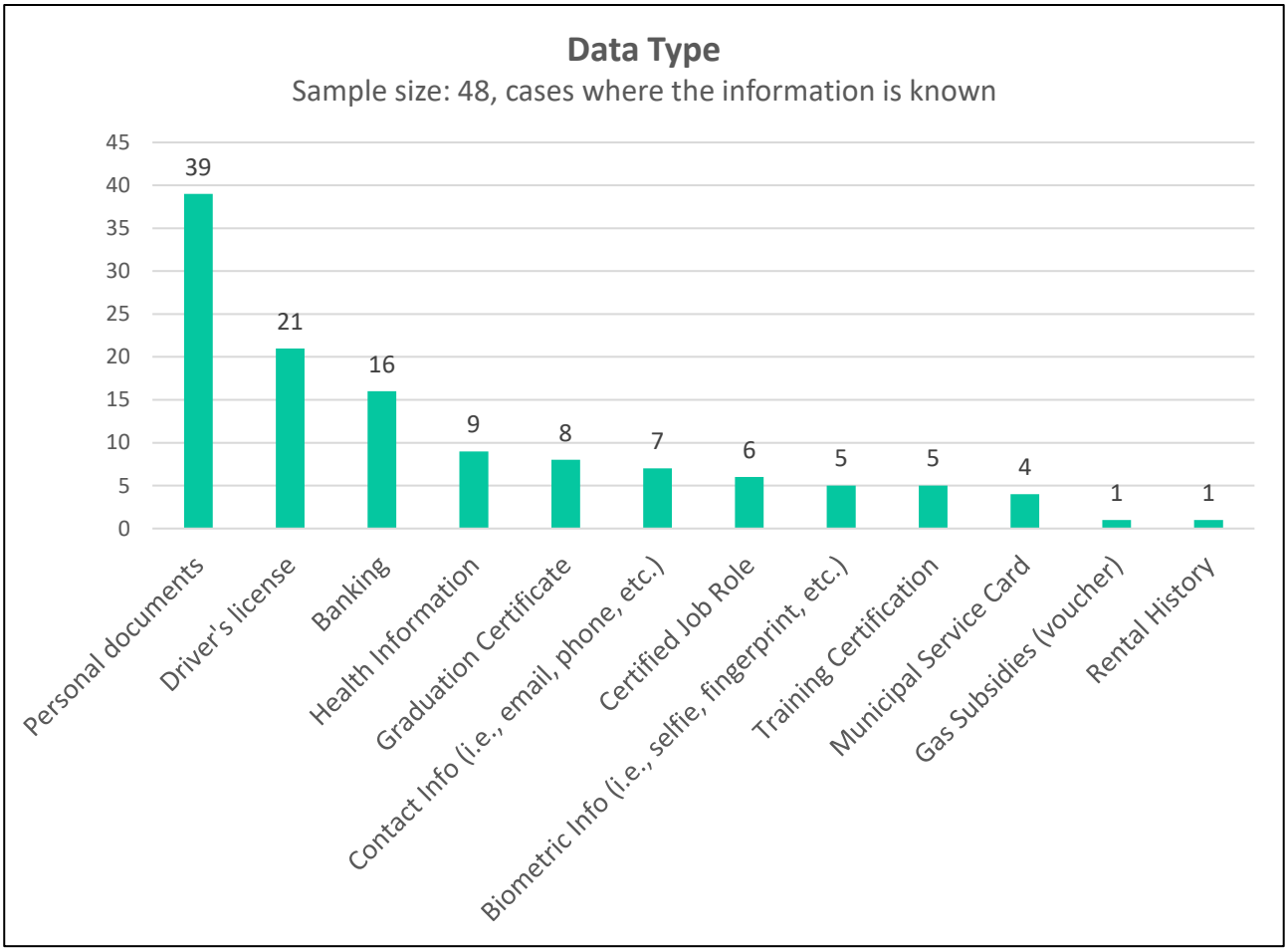

Figure 3. 17 Type of data associated with the identity profile (multiple types of data can be associated with the identity)

From *Figure 3.17* it can be understood that the most common data attributed to identity are personal documents (National ID, Passport, etc.) and driving license. In particular, there is a predominance of personal documents, which appear in 81% of cases. Then follow the data relating to the banking and healthcare context, two sectors that count numerous projects. There are also other types of data, but present in fewer cases. In theory, in SSI systems any attribute could be associated with identity. The results of the analysis on the practical cases show that the types of data actually attributed are a small set, compared to all those found. More specifically, 54% of cases allow to attribute only two or fewer types of data to the identity, effectively delineating a limited number of possible attributes for the identity. On the contrary, the projects that allow to attribute four or more data types to the identity are only 25% of the total. This means that, although few, there are some projects that allow users to associate a significantly broad set of attributes with the identity. Furthermore, the fact itself that there are practical cases, in which all these types of data are used, is interesting. It means, indeed, that such attributes are supported and usable in the SSI system. This is a significant step towards being able to associate any attribute to the identity profile. In addition, the possibility of attributing this wide array of data to identity, it is also a good sign as regards the interoperability of the model and its effective dissemination to all areas of daily life.

Without the possibility of associating certain attributes with identity, it would not be possible to operate in certain areas. For example, consider health data, which are often fundamental for healthcare projects. Precisely related to this, the results of the analysis that crosses the dimension of the data with that of the application areas yields interesting information. First, the various application areas do not involve specific requirements in terms of attributes associated with the identity. However, specific types of data, such as bank data or health data, are mainly used in the corresponding application areas, which in the example are finance and healthcare. Instead, crossing this perspective with others, such as active cases or cases in which institutions are present, no specific trend emerges. The results are aligned with those just presented.

The data described so far are always certified or linked to a physical document. Self-declaration is very rare and mainly concerns biometrics and contact information. It can therefore be said that with the exception of some basic data, the attributes of the identity are very reliable, as they are certified.

Among the most numerous types of data, there is that relating to banking data. This category includes a certain variety of attributes, which have been analyzed separately. The main results of the analysis are contained in *Figure 3.18*.

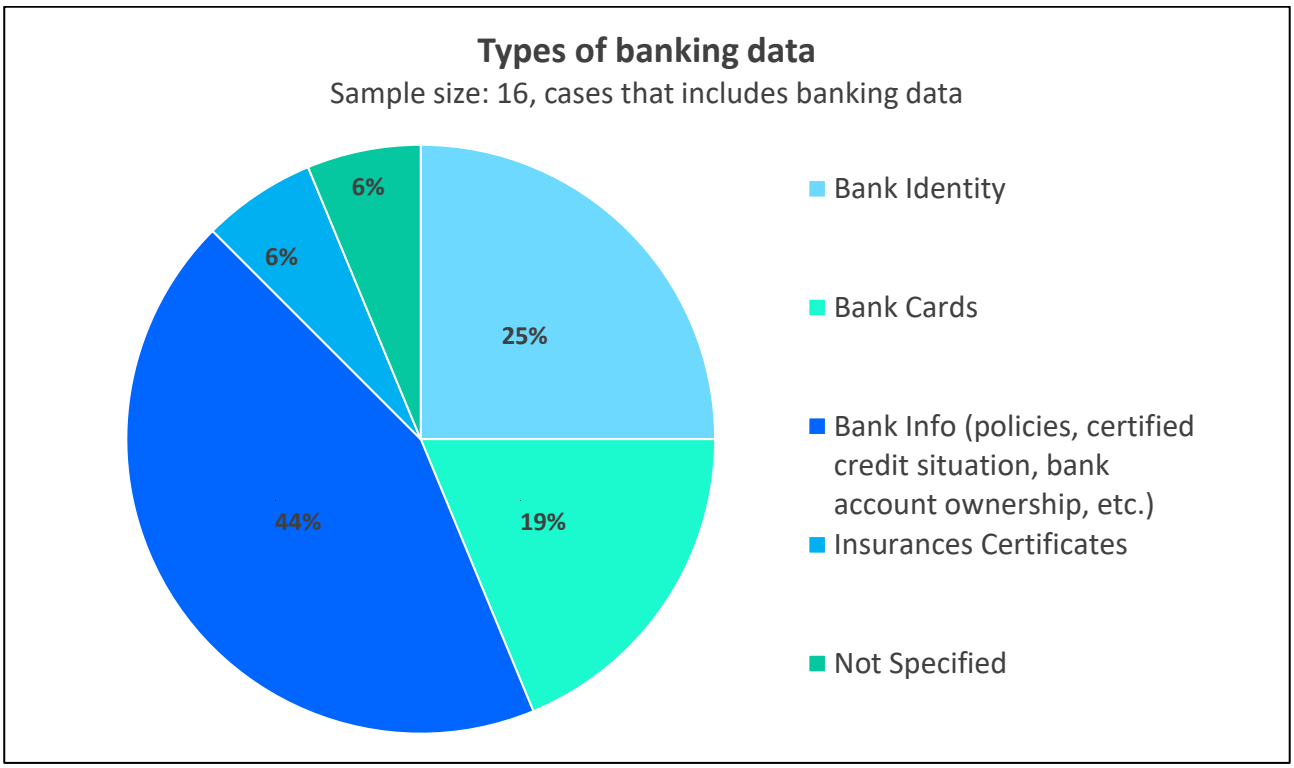

Figure 3. 18 Types of banking data

It should be noted that there is a tendency to certify some information, or the cards held by the user, in order to reuse them at a later time, sometimes for purposes other than banking. For example, the certified credit situation attribute is used in the Domi case, in the context of renting, to verify the reliability of a tenant. However, it does not replace other forms of recognition. This

feature is generally shared with most documents associated with the identity profile. They are saved for the purpose of reusing them at a later time. In the case of bank identities, instead, users can use them to identify themselves both in other financial companies and in other contexts. However, the latter are a minority of cases, and their functioning is not always clearly specified. To conclude, the aspect to be considered is that although specific and mainly used in the financial context, these kinds of attributes can be used in other contexts. This is also a valid observation for some other categories of specific dates, such as health data or data related to the world of work and education.

3.6 APPLICATION AREAS

3.6.1 Overall results

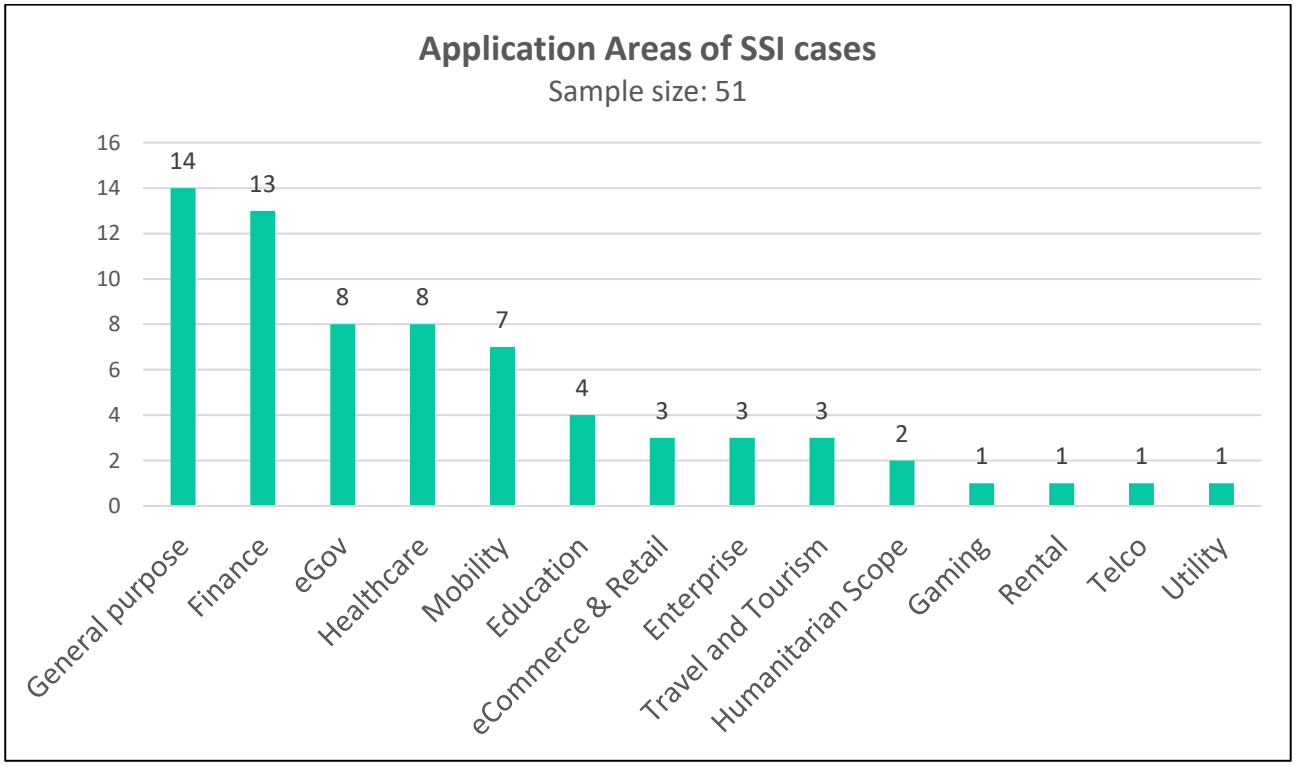

Figure 3. 19 Application areas of SSI cases (each case can have more than one application area)

The most popular projects are general purpose ones. Of all cases, 18 (35%) are of this type. However, of these 4 only in intention, as in practice they are limited to some application areas. Among the vertical applications fields, the most widespread, as can be seen from *Figure 3.19*, are finance, eGov, healthcare and mobility. There are projects in other areas too, but they are fewer in number. However, although still not very widespread in all these fields, the presence of projects in this wide

range of application areas is very positive for the spread of the SSI system. This means, indeed, that the identity model is suitable for multiple application areas, just as described at the theoretical level.

3.6.2 General Purpose projects

The results of the analysis on general purpose cases are based only on the 14 projects that are effectively of this type. Cases intended to be general purpose, but that are not in practice, have not been considered in this analysis.

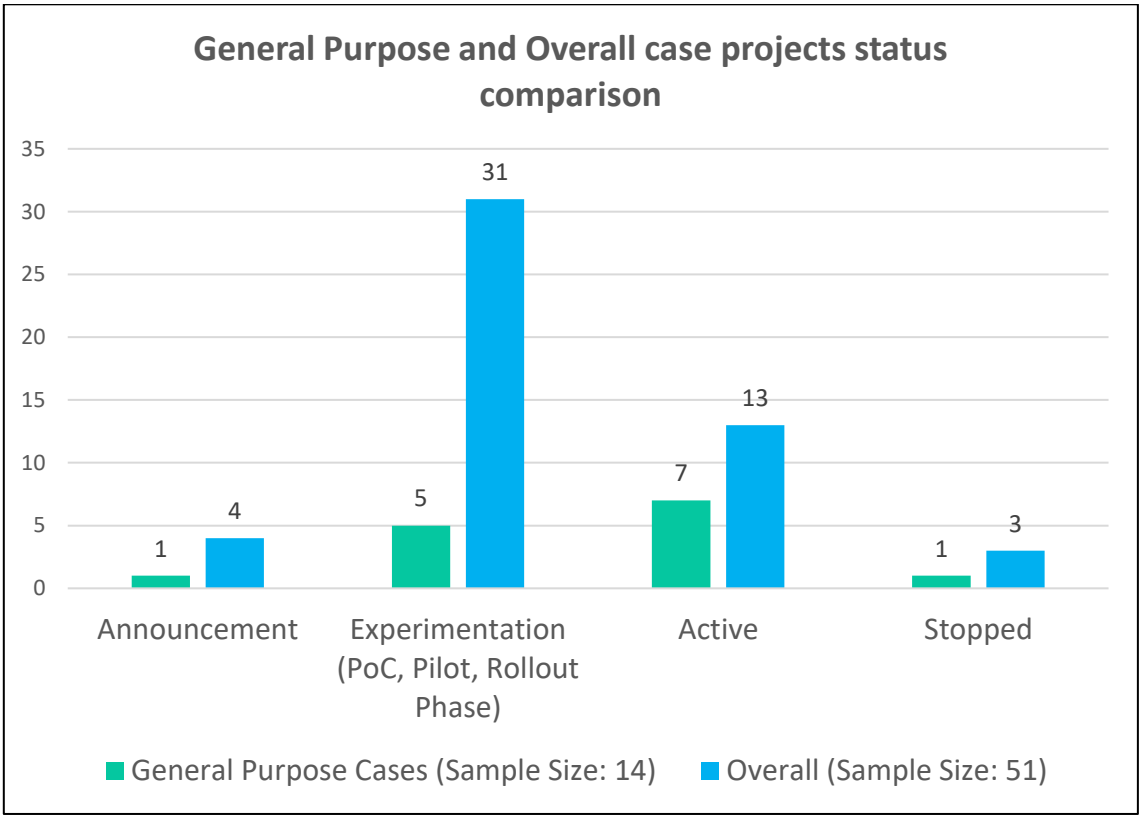

Figure 3. 20 General Purpose and Overall case projects status comparison

The outcomes of the analysis on the status of general purpose projects, summarized in *Figure 3.20*, returns very interesting data. 50% of these projects are active and they represent 54% of the total active projects. Although general purpose cases represent only 27% of total cases, they contribute to more than half of active projects. It is an encouraging sign, indeed, the idea behind the SSI model is that of an identity, which is not limited to specific sectors or niches, but which instead can have a wide variety of application areas. In this way it is possible to overcome the traditional paradigm and the user will be able to use a unique digital identity whenever necessary. In the theoretical description of the SSI model, this aspect is addressed through the interoperability principle.

As regards the geographical area, 10 different countries are represented (out of 19). As already mentioned, the only trend that emerges is that unlike the overall analysis, the US is the leading nation with 4 cases out of 14, a sign that there is attention to the development of an alternative and widely usable model in this country.

Looking at the actors involved, no differences emerge with the overall picture, there is a prevalence of the private sector with similar percentages. Going further into detail, when known, the ecosystems of general purpose projects are composed of a wide variety of actors, coming from multiple sectors. Among the types of partners most present there are obviously the technological ones. Outside of these, in the ecosystems are present local governments, partners from the industrial sector and to a lesser extent from finance, education and NGOs.

Similar speech for the protocols, there is not one prevalent and few with more than one case. An interesting fact is that the principle of interoperability in these projects tends to be respected more. However, the majority of cases still do not fully comply with the principle, due to the immaturity of the ecosystem and the limitations in terms of participants and in geographical terms.

Also, for the blockchain, its use and its characteristics are aligned with what emerged from the overall analysis.

Looking at data types, two interesting aspects emerge. The first is that all data types found are included at least once in one of the general-purpose projects, except the rental history, which however has only one case in all. The other aspect is that all general purpose cases include personal documents among their data, while for the other data there are percentages like those that emerged in the overall analysis. It can be implied that for a project of this type, users need an attribute that can be used in many contexts. Personal documents respond precisely to this need.

3.6.3 Application areas of projects with institutional bodies

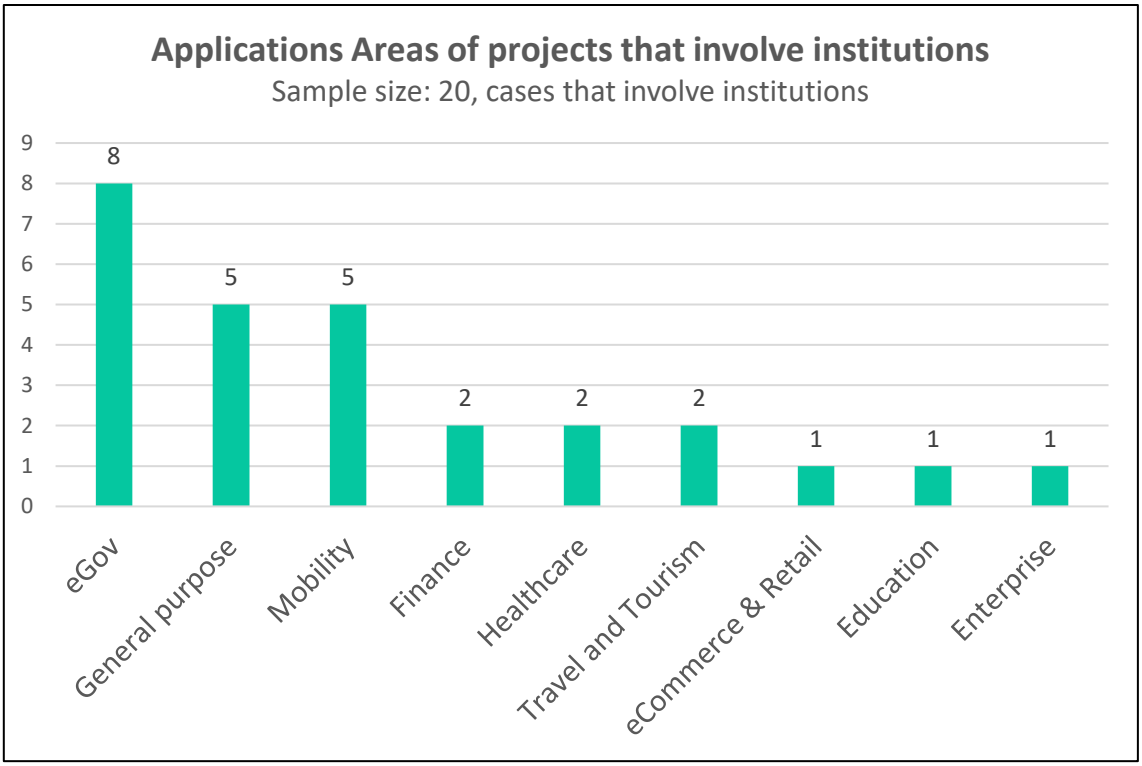

Figure 3. 21 Applications areas of projects that involve institutions (each case can have more than one application area)

Figure 3.21 shows that as expected, all projects under the eGov fields are carried out in an ecosystem that also includes the institutions themselves. The presence of numerous cases in the mobility sector is also not surprising given the strong interest of governments in Mobility as a Service (MaaS). General purpose cases also have a significant representation, in line with the fact that they are the most widespread cases overall. On the other hand, there are few cases relating to finance and healthcare in this context, which, instead, are rather popular overall. A sign that these sectors are strongly linked to private initiative.

3.7 DISCUSSION

The aim of the research was to study how the landscape of SSI type systems is configured at an international level. Thanks to this it is possible to understand how the theoretical concepts of the SSI model are applied in practice. To complete this quest multiple dimensions of analysis have been considered and combined with each other. The main evidence emerged and the main intercepted trends, divided according to the perspective to which they belong, are summarized in *Table 3.1*.

Perspective		Main findings
General Information	Geographical Distribution	➢ Europe is the continent with the most cases and different nations involved in a project. Africa is the least present, in line with reports from practitioner-oriented research centers. ➢ North America is strangely underrepresented in percentage, but the US has the largest number of active and general purpose projects.
	Project Status and Temporal Evolution	➢ Tangible projects (active and experimental) are the majority (86%). This means that the theoretical model is being translated into practical applications. ➢ The experimental part, 61% of total cases, is still prevalent over the rest. This confirms that the practical application of SSI is still in an embryonic stage. ➢ The percentage of active cases has increased over the years (with significant sample). This implies that solutions are increasingly ready to be used by end users.
	Main Actors	➢ The private sector is driving the model, being represented in all cases, and being present exclusively in most of them (61%). ➢ 85% of active cases include only private actors. From which it can be deduced that projects with institutions tend to stop earlier.
	Model Diffusion	➢ The number of users, when present (16% of cases), generally ranges from hundreds to a few thousand users involved. The adoption is very low, considering the potential, once again indicating the embryonic state of the practical application of the SSI model. The research on how to make it scalable is still ongoing. ➢ Success factors are to give the possibility to access as many services as possible with this identity and to be oriented to users' needs.
	Economic Sustainability	➢ Information on this aspect was found only in 33% of cases. Hence, it is evident that this is not one of the priorities and is a neglected factor not only by literature, but in some cases also on a practical level. ➢ The result of the analysis shows that there are no established business models. This is because it is an uncertain and strongly developing ecosystem (also considering associated technologies). ➢ An intercepted trend is that the user is generally not the paying party.
SSI Principles	SSI Protocols	➢ There are 23 protocols in all, of these only 5 are used for multiple use cases, but they cover 65% of the total cases. The lack of a precise international direction, also highlighted by

		the literature, implies that most cases try to develop their own protocol, especially when there is a geographical and cultural distance from the more established ones. ➢ The most promising protocols are Sovrin, Jolocom and W3C Standards, which alone account for 55% of total cases.
	SSI Principles	➢ All projects comply with the principles of Existence, Access, and Transparency. The principles of Control, Consent and Protection are also respected in the vast majority of cases. The most critical principles, on the other hand, are Persistence, Minimization and above all Portability and Interoperability. ➢ The difficulties encountered in several cases in fully respecting all ten principles confirm that the SSI model is still predominantly a theoretical model. ➢ However, the presence of cases and protocols capable of respecting all ten SSI principles and making technical choices motivated precisely by adapting to the theoretical dictates of the model, confirms that a transition is taking place towards the practical application of the SSI concept.
Technological Aspects	Blockchain	➢ The strong link that emerged in the literature between SSI and the blockchain is also confirmed in practice. 88% of cases use this technology. ➢ The blockchains used are mostly public (82%) and permissioned (56%). The first aspect is consistent with the theoretical model. Being permissioned, on the other hand, carries the risk of centralizing power in the hands of a small group of people. This is not aligned with the theoretical dictates of the SSI. However, in active projects the permissionless case (60%) prevails, which testifies to the applicability of this type of blockchain.
	Integration Technologies	➢ All cases, where the information is known, use integration technologies in the API & SDK category. The other protocols are used only for retrofitting with existing systems. API & SDK provide greater flexibility and facilitate the dissemination of the SSI model.
	Process Technologies	➢ In 79% of cases where information is available a wallet is used, while a mobile application is present in all cases. ➢ Mobile wallets (97% of the cases that use a wallet) and mobile applications have two advantages: portability and the ease with which they can be obtained. Indeed, nowadays many people have a smartphone on which to install them. This can greatly facilitate the access and diffusion of the SSI system. ➢ For mobile applications, there is a predominance in the use of QR Code. Biometric and PIN follow. The use of passwords is down compared to the traditional system.

User Perspective	Access/Use Credentials	➢ The most used factors are biometric and PIN. There is a sharp decline in the use of password and User ID, as in the case of process technologies. SSI projects mark a shift from the methods used in traditional systems towards more innovative technologies. These technologies are safer, but still very simple to use for the end users, improving their experience. ➢ MFA is compulsorily used only in 24% of cases. However, the fact that MFA is the direction of the future is evidenced by the fact that 77% of cases support it.
	Onboarding Procedures	➢ Online onboarding is the most common procedure (used in approximately 52% of cases). This can improve the user experience, as this type of onboarding saves time and costs. In addition, it facilitates access to identity for people residing in rural areas or with difficulty moving around. ➢ There are different types of online onboarding. The most common is the verification of documents uploaded by the user, which is facilitated using the blockchain. The documents verified most often are the most common ones, such as National ID, Passport and Driver's License. ➢ In person onboarding remains widespread, with only one case less than the online one. Especially in cases where institutions are involved in the project. In person onboarding is required in 80% of these cases. This limits the benefits described above, also considering that it is precisely the institutions that issue the most necessary documents for online onboarding.
Data		➢ The most common data, attributed to identity, are personal documents (National ID, Passport, etc.) and driving license. They appear respectively in 81% and 44% of cases. Below is the data relating to the banking and health sector. There are also other types of data, but present in fewer cases. ➢ The literature reports that theoretically in SSI systems any attribute could be associated with identity. However, the 54% of cases allow to attribute only two or fewer types of data to the identity, effectively delineating a limited number of possible attributes for the identity. On the contrary, the projects that allow to attribute four or more data types to the identity are only 25% of the total. So, although few, there are some projects that allow users to associate a significantly broad set of attributes with the identity. Furthermore, the fact that there are practical cases, in which all types of data found are used, means that such attributes are supported and usable in the SSI system. This is a significant step towards being able to associate any attribute to the identity profile. ➢ In all cases, the data described are certified or linked to a physical document (except for some basic information). This

		means that the attributes of the identity are very reliable, as they are certified. ➢ The results of the detailed analysis of banking data reveal that they are mainly used in the financial sector, however despite their specificity they can also be used in other application areas. This observation can be extended to other categories of specific data, such as health data or those relating to education and work.
Application Areas	Overall Results	➢ The most popular projects are general purpose ones (27% of total cases). ➢ Among the vertical applications fields, the most widespread, are finance, eGov, healthcare and mobility. There are projects in other areas too, but they are fewer in number. ➢ The presence of projects in the wide range of application areas found means that the SSI model is suitable for being used in any context, just as described at the theoretical level.
	General Purpose Projects	➢ 50% of general purpose projects are active. Although these projects represent only 27% of total cases, they count for 54% of the total active projects. This is an encouraging factor, aligned with what emerged from the literature, namely that the SSI is not limited to specific sectors or niches, but instead can have a wide variety of application areas. It is no coincidence that the principle of interoperability tends to be respected more in these cases (29% against 21% in the overall case). ➢ All general purpose cases include personal documents among the data attributed to identity. It can be implied that for a project of this type, users need an attribute that can be used in many contexts. Personal documents respond precisely to this need.
	Application areas of projects with institutional bodies	➢ All projects under the eGov fields are carried out in an ecosystem that also includes the institutions themselves. ➢ The presence of numerous cases in the mobility sector is due to the strong interest of governments in MaaS. ➢ There are few cases relating to finance and healthcare in this context, which, instead, are rather popular overall. A sign that these sectors are strongly linked to private initiative.

Table 3. 1 Recap of the main evidence and of the main trends intercepted following the empirical work

Chapter 4: Conclusions

This last chapter contains a sum-up of the main contributions of this thesis work. The chapter will conclude by explaining the limitations of this work and possible directions for future research.

4.1 RESULTS SUMMARY

The results described in the previous chapter confirm the presence of an emerging global ecosystem of Self-Sovereign Identity. This work was one of the first attempts to describe this ecosystem, trying to collect and analyze the data considered significant, following a thorough review of the existing literature.

As many as 47% of cases are in Europe, which is by far the continent most involved in the development of SSI-type solutions. Among the main reasons is the strong regulation in terms of privacy and protection of user data in that continent, with which this model is well suited. Asia and North America, which have fewer total cases, have more active projects, showing a strong practical interest in the model. Generally, these two continents are very active in the digital sector and an increase in interest in SSI would help accelerate the development of the ecosystem. Africa, on the other hand, is the least represented continent, confirming the difficulties of this territory in terms of digital identity, due to the lack of infrastructure and know-how. Looking at the status of the projects, the tangible projects represent a majority compared to the announcements and the stopped projects. This means that the theoretical interest is turning into concrete and practical applications of the SSI model. However, 61% of the projects are tests or pilots, while fully active projects are 25%. This is rather indicative of the embryonic stage in which the practical application of the SSI model still stands. The experimental part is still prevalent over the rest. Regarding this, it is interesting to note that the number of cases has grown steadily since 2017 until 2020, and above all, the percentage of active cases has grown steadily over the years. This is a sign that the proposed solutions are increasingly ready to be distributed on the market. Considering the actors present in the ecosystem, the private sector is driving the model, being represented in all cases, and being present exclusively in most of them (61%). The difference is even more marked if we consider the active cases, where 85% of cases rely only on the private sector. Similar results were obtained considering the SPs involved in the various projects. Lastly, the study of the diffusion of the model in numerical terms and of the economic model underlying the SSI system was carried out only at a qualitative level, as the novelty and the embryonic phase in which the model still finds itself prevented the finding of sufficient data to complete a quantitative analysis. The main observations regarding these two aspects are described in detail in the previous chapter.

Analyzing the 51 cases, 23 different SSI protocols emerged. Among these three stand out for the number of cases and attention to all aspects of the SSI model: Sovrin, Jolocom and the W3C Standards. These three alone account for 55% of total cases and cover multiple vertical application areas as well as general purpose projects. It is therefore possible to observe that the lack of a precise international direction implies that most projects try to develop their own protocol, especially when there is a geographical and cultural distance from the more established ones. On the other hand, it is also true that there are protocols that are establishing themselves as the main ones in the landscapes. Turning to the results of the analysis on the ten principles of the SSI model it is possible to say that this is a critical aspect for the proposed systems. Indeed, it is noted that some cases do not pay much consideration to these principles, still showing some confusion on the theoretical SSI model. In particular, four principles emerge as critical: Persistence, Minimization, Portability, and Interoperability. The latter two are not respected or are only partially respected in 67% and 78% of cases respectively. However, it should be specified that there are projects capable of satisfying all ten principles. Moreover, in many cases the non-compliance is due to the novelty of the model and the consequent limited environments. Further development of the ecosystem will lead to significant improvements in this regard.

As far as the technological aspects are concerned, the analysis made it possible to identify some consolidated trends and practices. First, the strong link between SSI and Blockchain, identified in the literature, is also confirmed on a practical level. Indeed, in 88% of cases the blockchain is used. These blockchains are mainly public and permissioned. This represents an obstacle to the full achievement of the potential of the SSI model. Indeed, the ideal case would be a permissionless platform, in order to avoid the presence of authorities that could centralize the control of the blockchain. From this point of view, the fact that among the 10 active cases that use blockchain technology, there is, instead, a prevalence of public and permissionless blockchains is to be considered much better and encouraging on the effective usability of this type of blockchain in operational cases. As for integration technologies, the results show a predominance of API & SDK, in the few cases where other technologies are used it is only to allow retrofitting with existing systems. The various benefits associated with the use of these technologies help to promote the diffusion of the SSI model. As for process technologies, in 79% of cases where information is available a wallet is used, while a mobile application is used in all cases. This can have a strong impact on the diffusion and access of the SSI model, indeed, they can be easily downloaded and used on a common smartphone. Looking at the characteristics of the mobile applications used, a trend emerges, also found in the aspects relating to the user's perspective, which is a wide use of innovative technologies and methods at the expense of traditional ones. In the specific case, methods such as PIN and Password are still quite widespread, but in sharp decline if compared to the traditional systems. At the same time emerging technologies such as QR Codes and Biometrics are used in most cases.

Similarly, the access methods analysis results show a decline in the use of User ID and Password, typical of traditional systems. Instead, the most used methods are biometrics and PIN. These technologies are safer, but still very simple to use for the end users, improving their experience. To further increase safety, MFA is supported in 77% of cases. Looking at the onboarding process, projects are homogenously distributed, with 17 cases of online onboarding and 16 of in person onboarding. This is an interesting result, indeed, online onboarding can reduce the time and costs required for the procedure. In addition, it would facilitate access to identity for people residing in rural areas or with difficulty moving around.

The way it was conceived, the SSI model foresees that any type of attribute can be associated to the identity. This is not completely respected in practice, indeed, 54% of cases allow to attribute only two or fewer types of data to the identity. It can be said that in most cases the identity profile has a limited set of attributes. There are, however, cases that allow to attribute multiple attributes to the identity. In addition, more generally several types of data were found during the analysis of the projects. This means that this wide set of attributes is supported and usable in the SSI system, thus being a solid basis for the future achievement of the goal. Furthermore, the analysis on the data showed that they are always certified or linked to a physical document, which allows them to be considered reliable. Self-declaration is very rare and mainly concerns biometrics and contact information.

Lastly, considering the application areas, the most common projects are general purpose ones (27%). While as regards the vertical applications fields, the most widespread are finance, eGov, healthcare and mobility. There are projects in other areas too, but they are fewer in number. This means that the SSI identity model is suitable for multiple application areas, just as described at the theoretical level.

4.2 LIMITATIONS AND FUTURE RESEARCH

The methodology presented in Chapter 2 has been designed and reviewed during the research to obtain a research process as rigorous and replicable as possible, with the ultimate goal of creating a reliable source of information on the Self-Sovereign Identity ecosystem. However, considering the limitation of resources and the inevitable research bias, some limitations remain regarding data collection and the relative analysis:

- **Data Sources**: the research for existing SSI cases involved the use of two different types of data sources: literature and specialized websites in the sector. In this way it was possible to cover heterogeneous research areas. However, it is not certain that all SSI cases have been found, precisely because for example not included in one of the data sources. In addition, a

further limitation derives from the use of the English language to carry out the research. This may have caused an imbalance among cases developed in countries where this language is widespread and used at the expense of others, such as China or Russia.

- **Extraction Process**: before carrying out the research on the specialized websites in the sector, a series of specific keywords to be used for the search was defined. This set was built in a specific way also to avoid having vague or uninteresting results for the research. However, this may have led to the exclusion of valid cases with a vague or inaccurate description.
- **Data Integration**: the dataset extracted from the data sources was integrated with a series of information relating to specific variables, considered relevant following the literature review process. The author's judgment may have diverted the result of the process in considering relevant information or not. Moreover, the objectivity of the information present on institutional websites could be questioned.
- **Scientific Material**: the novelty of the argument has limited the number of documents and papers available, being the Self-Sovereign Identity topic in its early phases.

Given its originality in the object of study, this thesis could lay the foundation for further studies in the field of digital identity. In continuity with this research work, several directions for future research could be outlined:

- The database could be periodically updated, considering changes in the ecosystem. In this way it would be possible to make observations on the progress and diffusion of the SSI model.
- Some successful cases and protocols were presented in the research. Through further research it is possible to verify their status and whether any new success cases are inspired by them. In this way it would be possible to formalize the success factors for the SSI model.
- During the work it was mentioned that only with the maturity of the SSI model it will be possible to know more precise information on some variables. These include, for example, the diffusion of the model, the economic sustainability, or the compliance with the principles of interoperability and portability. One or more of these areas can be the targeted focus for future research, in order to answer these open questions.

Bibliography

Ali O., Jaradat A., Kulakli A., Abuhalimeh A., 2021, "*A comparative study: Blockchain technology utilization benefits, challenges and functionalities*", IEEE Access, Vol. 9, p. 12730-12749, DOI: 10.1109/ACCESS.2021.3050241

Allison A., Currall J., Moss M., & Stuart S., 2005, "*Digital identity matters*", Journal of the American Society for Information Science and Technology, Vol. 56, p. 364-372, https://doi.org/10.1002/asi.20112

Andermatt K., Göldin R., 2018, "*Introducing an Electronic Identity: The Co-design Approach in the Canton of Schaffhausen*", Swiss Yearbook of Administrative Sciences, Vol. 9, p. 41-50, DOI: 10.5334/ssas.122.

Arner D.W., Zetzsche D.A., Buckley R.P., Barberis J.N., 2019, "*The Identity Challenge in Finance: From Analogue Identity to Digitized Identification to Digital KYC Utilities*", European Business Organization Law Review, Vol. 20, p. 55-80, DOI: 10.1007/s40804-019-00135-1

Atick J., 2016, "*Digital identity: the essential guide*", ID4Africa Identity Forum, available at: https://www.id4africa.com/main/files/Digital_Identity_The_Essential_Guide.pdf

Bandara E., Liang X., Foytik P., Shetty S., Hall C., Bowden D., Ranasinghe N., De Zoysa K., 2021, "*A blockchain empowered and privacy preserving digital contact tracing platform*", Information Processing & Management, Vol. 58, https://doi.org/10.1016/j.ipm.2021.102572

Beduschi A., 2019, "*Digital identity: Contemporary challenges for data protection, privacy and non-discrimination rights*", Big Data & Society, Vol.6, https://doi.org/10.1177/2053951719855091

Bernal Bernabe J., Canovas J. L., Hernandez-Ramos J. L., Torres Moreno R., Skarmeta A., 2019, "*Privacy-Preserving Solutions for Blockchain: Review and Challenges*", IEEE Access, Vol. 7, p. 164908-164940, DOI: 10.1109/ACCESS.2019.2950872.

Bertino E., Bhargav-Spantzel A., Elliott S.J., Modi S., Squicciarini A.C., Young M., 2007, "*Privacy preserving multi-factor authentication with biometrics*", Journal of Computer Security, Vol. 15, p. 529-560, DOI:10.3233/JCS-2007-15503

Blockchain and Distributed Ledger Observatory of Politecnico di Milano, 2020, "*Self-Sovereign Identity e Blockchain: binomio vincente?*", research report

Bochem A., Leiding B., 2021, "*Rechained: Sybil-Resistant Distributed Identities for the Internet of Things and Mobile Ad Hoc Networks",* Sensors, Vol. 21, https://doi.org/10.3390/s21093257

Cameron A., Grewe O., 2022, "*An Overview of the Digital Identity Lifecycle (v2)*", IDPro Body of Knowledge, https://doi.org/10.55621/idpro.31

Camp L. J., 2004, "*Digital identity*", IEEE Technology and Society Magazine, Vol. 23, p. 34-41, DOI:10.1109/MTAS.2004.1337889

Chadwick D. W., 2009, "*Federated identity management*", Foundations of Security Analysis and Design V, Springer, p. 96-120, DOI:10.1007/978-3-642-03829-7_3

Cheesman M., 2022, "*Self-Sovereignty for Refugees? The Contested Horizons of Digital Identity*", Geopolitics, Vol. 27, p. 134-159, DOI: 10.1080/14650045.2020.1823836

Clark J., Gelb A., 2013, "*Identification for Development: The Biometrics Revolution*", Center for Global Development Working Paper No. 315, DOI:10.2139/ssrn.2226594

Cocco L., Tonelli R., Marchesi M., 2021, "*Blockchain and Self Sovereign Identity to Support Quality in the Food Supply Chain. Future Internet*", Vol. 13, p. 301, https://doi.org/10.3390/fi13120301

Čučko Š., Turkanović M., 2021, "*Decentralized and Self-Sovereign Identity: Systematic Mapping Study*", IEEE Access, Vol. 9, p. 139009-139027, DOI: 10.1109/ACCESS.2021.3117588.

De Diego S., Regueiro C., Maciá-Fernández G., 2021, "*Enabling Identity for the IoT-as-a-Service Business Model*", IEEE Access, Vol. 9, p. 159965-159975, DOI: 10.1109/ACCESS.2021.3131012

Der U., Jähnichen S., Sürmeli J., 2017, "*Self-sovereign Identity - Opportunities and Challenges for the Digital Revolution*", Computer Science, https://doi.org/10.48550/arXiv.1712.01767

Di Pierro M., 2017, "*What is the blockchain?*", Computing in Science & Engineering, Vol. 19, p. 92-95, DOI:10.1109/MCSE.2017.3421554

Di Sarno P., 2020, "*An international census of startups in the digital identity and access management sector*", http://hdl.handle.net/10589/167535

Dib O., Toumi K., 2020, "*Decentralized Identity Systems: Architecture, Challenges, Solutions and Future Directions*", Annals of Emerging Technologies in Computing, Vol. 4 N. 5, p. 19-40, DOI: 10.33166/AETiC.2020.05.002

Digital Identity Observatory, Politecnico di Milano, 2020, "*L'identità digitale: definizioni e scenario generale*", research report

Digital Identity Observatory, Politecnico di Milano, 2020, "*Alla ricerca dell'identità… digitale*"

Dixon P., 2017, "*A Failure to "Do No Harm" -- India's Aadhaar biometric ID program and its inability to protect privacy in relation to measures in Europe and the US*", Health and technology, Vol. 7, p. 539-567, https://doi.org/10.1007/s12553-017-0202-6

Dunphy P., Petitcolas F. A. P., 2018, "*A First Look at Identity Management Schemes on the Blockchain*", IEEE Security & Privacy, Vol. 16, No. 4, p. 20-29, DOI: 10.1109/MSP.2018.3111247

Estevam G., Palma L., Silva L., Martina J., Vigil M., 2021, "*Accurate and decentralized timestamping using smart contracts on the Ethereum blockchain*", Information Processing & Management, Vol. 58, https://doi.org/10.1016/j.ipm.2020.102471

Ferdous M. S., Chowdhury F., Alassafi M. O., 2019, "*In Search of Self-Sovereign Identity Leveraging Blockchain Technology*", IEEE Access, Vol. 7, p. 103059-103079, DOI: 10.1109/ACCESS.2019.2931173.

Fridgen G., Guggenmos F., Lockl J., Rieger A., 2018, "*Challenges and opportunities of blockchain-based platformization of digital identities in the public sector*", Workshop on Platformization in the Public Sector, 26th European Conference on Information Systems (ECIS)

Gasser U., Palfrey J. G., 2007, "*Case Study: Digital Identity Interoperability and eInnovation*", Berkman Center Research Publication, No. 2007-11, http://dx.doi.org/10.2139/ssrn.1070061

Gelb A., Metz A. D., 2018, "*Identification Revolution: Can Digital ID be Harnessed for Development?*", Center for Global Development, p. 5-9, 23-30, 59-91

Giannopoulou A., Wang F., 2021, "*Self-sovereign identity*", Internet Policy Review, Vol. 10, https://doi.org/10.14763/2021.2.1550

Gipp B., Breitinger C., Meuschke N., Beel J., 2017, "*CryptSubmit: Introducing Securely Timestamped Manuscript Submission and Peer Review Feedback Using the Blockchain*", 2017 ACM/IEEE Joint Conference on Digital Libraries, p. 1-4, DOI: 10.1109/JCDL.2017.7991588.

Goldwasser S., Micali S., Rackoff C., 1989, "*The knowledge complexity of interactive proof systems*", SIAM Journal on Computing, Vol. 18, No. 1, p. 186-208

Grüner A., Mühle A., Meinel C., 2021, "*ATIB: Design and Evaluation of an Architecture for Brokered Self-Sovereign Identity Integration and Trust-Enhancing Attribute Aggregation for Service Provider*", IEEE Access, Vol. 9, p. 138553-138570, DOI: 10.1109/ACCESS.2021.3116095.

Gstrein O.J., Yap E., Zwitter A.J., 2020, "*Digital Identity and the Blockchain: Universal Identity Management and the Concept of the "Self-Sovereign" Individual*", Front. Blockchain, https://doi.org/10.3389/fbloc.2020.00026

Hasan H., Salah K., Jayaraman R., Arshad J., Yaqoob I., Omar M., Ellahham S., 2020, "*Blockchain-based Solution for COVID-19 Digital Medical Passports and Immunity Certificates*", IEEE Access, vol. 8, p. 222093-222108, DOI: 10.1109/ACCESS.2020.3043350

Houtan B., Hafid A. S., Makrakis D., 2020, "*A Survey on Blockchain-Based Self-Sovereign Patient Identity in Healthcare*", IEEE Access, Vol. 8, p. 90478-90494, DOI: 10.1109/ACCESS.2020.2994090

Hughes J., Maler E., 2005, "*Security assertion markup language (saml) v2. 0 technical overview*". OASIS SSTC Working Draft sstc-saml-tech-overview-2.0-draft-08, available at: http://docs.oasis-open.org/security/saml/Post2.0/sstc-saml-tech-overview-2.0-cd-02.pdf

Ishmaev G., 2020, "*Sovereignty, privacy, and ethics in blockchain-based identity management systems*", Ethics and Information Technology, Vol. 23, p. 239-252, https://doi.org/10.1007/s10676-020-09563-x

Jacobovitz O., 2016, "*Blockchain for identity management*", The Lynne and William Frankel Center for Computer Science Department of Computer Science, Ben-Gurion University, Beer Sheva 1

Janssen M., Brous P., Estevez E., Barbosa L., Janowski T., 2020, "*Data governance: Organizing data for trustworthy Artificial Intelligence*", Government Information Quarterly, Vol. 37, DOI: 10.1016/j.giq.2020.101493

Jøsang A., Pope S., 2005, "*User centric identity management*", Proceedings of the AusCERT Asia Pacific Information Technology Security Conference, p. 77

Körner M. F., Sedlmeir J., Weibelzahl M., Fridgen G., Heine M., Neumann C., 2022, "*Systemic risks in electricity systems: A perspective on the potential of digital technologies*", Energy Policy, Vol. 164, https://doi.org/10.1016/j.enpol.2022.112901

Laatikainen G., Kolehmainen T., Abrahamsson P., 2021, "*Self-sovereign identity ecosystems: benefits and challenges*", Scandinavian Conference on Information Systems, Association for Information Systems, available at: https://aisel.aisnet.org/scis2021/10/

Lesavre L., Varin P., Mell P., Davidson M., Shook J., 2020, "*A taxonomic approach to understanding emerging blockchain identity management systems*", National Institute of Standards and Technology, https://doi.org/10.6028/NIST.CSWP.01142020

Li X., Wu X., Pei X., Yao Z., 2019, "*Tokenization: Open Asset Protocol on Blockchain*", Proceedings of the 2019 IEEE 2nd International Conference on Information and Computer Technologies, p. 204-209, DOI: 10.1109/INFOCT.2019.8711021

Liu Y., He D., Obaidat M. S., Kumar N., Khan M. K., Choo K. K. R., 2020, "*Blockchain-based identity management systems: A review*", Journal of network and computer applications, Vol. 166, p. 102731, https://doi.org/10.1016/j.jnca.2020.102731

Liu Y., Lu Q., Zhu C., Yu Q., 2021, "*A blockchain-based platform architecture for multimedia data management*", Multimedia Tools and Applications, Vol. 80, p. 30707-30723, https://doi.org/10.1007/s11042-021-10558-z

Loffreto D., 2012, "*What is 'Sovereign Source Authority'?*", Web Article, The Moxy Tongu, available at: https://www.moxytongue.com/2012/02/what-is-sovereign-source-authority.html

MacKenzie D., 2019, "*Pick a nonce and try a hash*", London Review of Books, Vol. 41(8), p. 35-38, https://www.lrb.co.uk/v41/n08/donald-mackenzie/pick-a-nonce-and-try-a-hash

McKinsey Global Institute, 2019, "*Digital Identification a key to inclusive growth*", available at: https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Digital%20identification%20A%20key%20to%20inclusive%20growth/MGI-Digital-identification-Report.ashx

Mohammadzadeh N., Dorri Nogoorani S., Munoz-Tapia J. L., 2021, "Decentralized Factoring for Self-Sovereign Identities", Electronics, Vol. 10, 1467. https://doi.org/10.3390/electronics10121467

Mühle A., Grüner A., Gayvoronskaya T., Meinel C., 2018, "*A survey on essential components of a self-sovereign identity*", Computer Science Review, Vol. 30, p. 80-86, https://doi.org/10.1016/j.cosrev.2018.10.002.

Muralidharan K., Niehaus P., Sukhtankar S., 2020, "*Balancing corruption and exclusion: Incorporating Aadhaar into PDS*", Ideas for India, available at: https://www.ideasforindia.in/topics/poverty-inequality/balancing-corruption-and-exclusion-incorporating-aadhaar-into-pds.html

National Institute of Standards and Technology, U.S. Department of Commerce, 2017, "*Digital Identity Guidelines*" Available at: http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-63-3.pdf.

Nofer M., Gomber P., Hinz O., Schiereck D., 2017, *"Blockchain"*, Business & Information Systems Engineering, Vol. 59(3), p. 183-187, https://doi.org/10.1007/s12599-017-0467-3

Nyst C., Makin P., Pannifer S., Whitley E. A., 2016, "*Digital identity: Issue analysis: executive summary*", Consult Hyperion, Guildford

Oxford University, 2022, Oxford Advanced Learner's Dictionary. Available at: https://www.oxfordlearnersdictionaries.com/definition/english/entity?q=entity

Ometov A., Bezzateev S., Mäkitalo N., Andreev S., Mikkonen T., Koucheryavy Y., 2018, "*Multi-Factor Authentication: A Survey*", Cryptography, Vol. 2(1), https://doi.org/10.3390/cryptography2010001

Panait A. E., Olimid R. F., Stefanescu A., 2020, "*Identity Management on Blockchain--Privacy and Security Aspects*", https://doi.org/10.48550/arXiv.2004.13107

Pennino D., Pizzonia M., Vitaletti A., Zecchini M., 2021, "*Efficient Certification of Endpoint Control on Blockchain*", IEEE Access, Vol. 9, p. 133309-133334, DOI: 10.1109/ACCESS.2021.3115343

Pöhn D., Hommel W., 2020, "*An overview of limitations and approaches in identity management*", Proceedings of the 15th International Conference on Availability, Reliability and Security, p. 1-10

Pöhn D., Grabatin M., Hommel W., 2021, "*eID and Self-Sovereign Identity Usage: An Overview*", Electronics, Vol. 10, https://doi.org/10.3390/electronics10222811

Rosen J., 2011, "*The right to be forgotten*", Stan. L. Rev. Online, Vol. 64, available at: https://www.stanfordlawreview.org/online/privacy-paradox-the-right-to-be-forgotten/

Rundle M. C., Trevithick P., 2007, "*Interoperability in the New Digital Identity Infrastructure*", SSRN Electronic Journal, http://dx.doi.org/10.2139/ssrn.962701

Sakimura N., Bradley J., Jones M., De Medeiros B., Mortimore C., 2014, "*Openid connect core 1.0*", The OpenID Foundation, available at: https://openid.net/specs/openid-connect-core-1_0-final.html

Saleh F., 2021, "*Blockchain without waste: Proof-of-stake*", The Review of financial studies, Vol. 34, p. 1156-1190, https://doi.org/10.1093/rfs/hhaa075

Salleras X., Daza V., 2020, "*SANS: Self-Sovereign Authentication for Network Slices*", Security and Communication Networks, Vol. 2020, https://doi.org/10.1155/2020/8823573

Salleras X., Daza V., 2021, "*ZPiE: Zero-Knowledge Proofs in Embedded Systems*", Mathematics, Vol. 9 (20), https://doi.org/10.3390/math9202569

Sarmah S., 2018, "*Understanding Blockchain Technology*", Computer Science and Engineering, Vol. 8, p. 23-29, DOI: 10.5923/j.computer.20180802.02

Satybaldy A., Nowostawski M., Ellingsen J., (2019), "*Self-sovereign identity systems*", IFIP International Summer School on Privacy and Identity Management, p. 447-461, Springer, Cham.

Schlatt V., Sedlmeir J., Feulner S., Urbach N., 2021, "*Designing a framework for digital KYC processes built on blockchain-based self-sovereign identity*", Information & Management, https://doi.org/10.1016/j.im.2021.103553

Sedlmeir J., Smethurst R., Rieger A., Fridgen G., 2021, "*Digital Identities and Verifiable Credentials*", Business & Information Systems Engineering, Vol. 63, p. 603-613, https://doi.org/10.1007/s12599-021-00722-y

Sedlmeir J., Lautenschlager J., Fridgen G., Urbach N., 2022, "*The transparency challenge of blockchain in organizations*", Electron Markets, https://doi.org/10.1007/s12525-022-00536-0

Segendorf B., 2014, "*What is Bitcoin?*", Sveriges Riksbank Economic Review, p. 71-87, available at: http://archive.riksbank.se/Documents/Rapporter/POV/2014/2014_2/rap_pov_1400918_eng.pdf

Sherman A. T., Javani F., Zhang H., Golaszewski E., 2019, "*On the Origins and Variations of Blockchain Technologies*", IEEE Security & Privacy, Vol. 17, p. 72-77, DOI: 10.1109/MSEC.2019.2893730

Soltani R., Nguyen U. T., An A., 2020, "*Decentralized and Privacy-Preserving Key Management Model*", Computers and Communications (ISNCC), p. 1-7, DOI: 10.1109/ISNCC49221.2020.9297294.

Soltani R., Nguyen U.T., An A., 2021, "*A Survey of Self-Sovereign Identity Ecosystem*", Security and Communication Networks, https://doi.org/10.1155/2021/8873429

Sovrin, 2016, "*The Inevitable Rise of Self-Sovereign Identity*", White Paper, available at: https://sovrin.org/wp-content/uploads/2018/03/The-Inevitable-Rise-of-Self-Sovereign-Identity.pdf

Sporny M., Longley D., Chadwick D., 2019, "*Verifiable credentials data model 1.0: Expressing verifiable information on the web*", available at: https://www.w3.org/TR/vc-data-model/, accessed August 2022

Sullivan C., 2012, "*Digital identity and mistake*", International Journal of Law and Information Technology, Vol. 20, No. 3, p. 224-229, DOI:10.1093/ijlit/eas015

Takemiya M., Vanieiev B., 2018, "*Sora Identity: Secure, Digital Identity on the Blockchain*", 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), p. 582-587, DOI: 10.1109/COMPSAC.2018.10299

USAID, 2017, "*Identity in a digital age: infrastructure for inclusive development*", report, available at: https://www.usaid.gov/sites/default/files/documents/15396/IDENTITY_IN_A_DIGITAL_AGE.pdf

Venkatraman S., Parvin S., 2022, "*Developing an IoT Identity Management System Using Blockchain*", Systems 2022, Vol. 10, https://doi.org/10.3390/systems10020039

Weitzberg K., 2020, "*Biometrics, race making, and white exceptionalism: The controversy over universal fingerprinting in Kenya*", The Journal of African History, Vol. 61, p. 23-43, https://doi.org/10.1017/S002185372000002X

World Bank Group, GSMA and Secure Identity Alliance, 2016, "*Digital Identity: Towards Shared Principles for Public and Private Sector Cooperation*", discussion paper, available at: https://documents1.worldbank.org/curated/en/600821469220400272/pdf/107201-WP-PUBLIC-WB-GSMA-SIADigitalIdentity-WEB.pdf

World Bank Group, 2017, "*Principles on identification for sustainable development*", Document, available at: https://documents1.worldbank.org/curated/en/213581486378184357/pdf/Principles-on-Identification-for-Sustainable-Development-Toward-the-Digital-Age.pdf

World Bank Group, 2018, "*Technology Landscape for Digital Identification*", Report, available at: https://documents1.worldbank.org/curated/en/199411519691370495/Technology-Landscape-for-Digital-Identification.pdf

World Economic Forum, 2016, “*A Blueprint for Digital Identity the Role of Financial Institutions in Building Digital Identity*”, World Economic Forum, Future of Financial Services Series, p.41, available at: https://www3.weforum.org/docs/WEF_A_Blueprint_for_Digital_Identity.pdf

World Economic Forum, 2019, “*The Global Risks Report*”, available at: https://www.weforum.org/reports/the-global-risks-report-2019/

Zheng Z., Xie S., Dai H., Chen X., Wang H., 2017, "*An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends*", 2017 IEEE International Congress on Big Data, p. 557-564, DOI: 10.1109/BigDataCongress.2017.85.

List of Figures

List of Tables

List of Abbreviations

API: Application Programming Interface

CA: Certification Authority

CRUD: Creating, Reading, Updating, and Deleting

DDO: DID Document

DID: Decentralized Identifier

DNS: Domain Name Server

DPKI: Decentralized Public Key Infrastructure

ICANN: Internet Corporation for Assigned Names and Numbers

IdaaS: Identity as a Service

IdP: Identity Provider

IPS: Interactive Proof Systems

KYC: Know Your Customer

LoA: Level of Assurance

MaaS: Mobility as a Service

MFA: Multi-Factor Authentication

NIPS: Non-Interactive Proof System

OAuth: Open Authentication

OIDC: OpenID Connect

PKI: Public Key Infrastructure

PoC: Proof of Concept

PoS: Proof-of-Stake

PoW: Proof-of-Work

SAML: Security Assertion Markup Language

SDK: Software Development Kit

SP: Service Provider

SPID: Sistema Pubblico di Identità Digitale

SSI: Self-Sovereign Identity

VCs: Verifiable Credentials

VP: Verifiable Presentation

W3C: World Wide Web Consortium

ZKP: Zero Knowledge Proof

zk-SNARK: Zero-Knowledge Succinct and Noninteractive ARguments of Knowledge

Acknowledgements

A conclusione del lavoro, volevo dedicare qualche riga a tutte le persone che mi sono state vicino durante questo percorso di crescita personale e professionale.

Volevo ringraziare Luca Gastaldi per avermi dato la possibilità di approfondire il tema scelto per questa tesi.

Un ringraziamento particolare e sentito a Clarissa Falcone, Diletta Villa e Giorgia Paola Dragoni per avermi supportato durante tutto il lavoro di ricerca e di stesura della tesi, fornendomi sempre preziosi consigli.

Un ringraziamento speciale ad Alessandra, alla mia famiglia e alle amiche e amici dell'Olimpo, che mi hanno supportato e sopportato in tutti questi anni, condividendo con me le gioie e aiutandomi nei periodi più difficili della mia esperienza universitaria.

Infine, un grazie a tutti i miei amici e ai miei compagni di studi per essermi sempre stati vicini e avermi regalato momenti di allegria e tranquillità.

Grazie di cuore a tutti voi.