University of Padua

Department of Mathematics "Tullio Levi-Civita"

Bachelor's Degree in Computer Science

Owning your data through Self-Sovereign Identity: agents implementation for Verifiable Credentials interaction

Bachelor thesis

Supervisor
Prof. Alessandro Brighente

Co-Supervisors
Prof. Mauro Conti
Dott. Mattia Zago

Graduating
Matteo Casonato

Academic Year 2021-2022

Matteo Casonato: *Owning your data through Self-Sovereign Identity: agents implementation for Verifiable Credentials interaction,* Bachelor thesis, © September 2022.

Abstract

Nowadays, most of our data is owned by private companies, and everyone knows everything about us because privacy online is not well preserved. Imagining a world different from this is difficult, but things can change thanks to Self-Sovereign Identity (SSI). SSI approach aims to bring credentials back to the actual owners, the people. This is possible through cryptography and secure authentication layers (e.g., OAuth, OpenIDConnect). The developed product embraces this philosophy and offers a solution where the users are the holders, issuers, or verifiers of Verifiable Credentials (VCs). Specifically, will be developed software agents who create, issue, verify, modify or even revoke the credentials, leveraging an SSI Kit.

In this thesis, we propose a methodology to merge SSI off-chain (i.e., outside the blockchain) operations with on-chain smart contracts. In particular, the job has been divided into three macro stages: firstly, has been done a deep dive into the SSI technology, studying all of its primitives and analyzing the problem; secondly, has been developed a Software Development Kit (SDK), which enabled us to dialog with an SSI Kit (off-chain logic); in the meantime, my friend and co-worker Matteo Midena developed the smart contracts (on-chain logic); finally, off-chain and on-chain solutions has been merged in a proof of concept web application.

“If you always do what you’ve always done,
you’ll always get what you’ve always got.”

— Henry Ford

Acknowledgments

First of all, I would like to thank the people who helped me during the writing of this paper: my supervisor Dott. Alessandro Brighente and my co-supervisors Prof. Mauro Conti and Dott. Mattia Zago.

Also, I want to thank my parents, who have always supported me, and never stopped me from doing anything I truly wanted.

Finally, I thank my friends, who have eased these sometimes intense but very satisfying years.

Padova, September 2022 Matteo Casonato

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter will introduce the problem: what we are analyzing, why this problem exists, how it is defined, and how it can be resolved. Also we present the company, the internship, and the work methodology.

1.1 The problem

As stated in the abstract, the main problem is preserving the ownership of people's data. In order to achieve this objective, we will pass through problems like interoperability, privacy safeguarding, law compliance, security, and others.

Assuming we can create a system where people hold their credentials (called Verifiable Credentials):

1. **Interoperability**: how can these credentials be shown to and verified in the same manner by different actors?
2. **Privacy**: can we demonstrate something without revealing it, preserving our privacy this way?
3. **Law compliance**: is it possible to save on blockchain people's data, or are we going against specific privacy laws?
4. **Security**: are credentials susceptible to attacks from hackers trying to steal our data?

Thanks to Self-Sovereign Identity, we can give a positive answer to all of these questions, but as is often the case, we have to deal with compromises.

Worldwide scenario. The current situation is clear. Every time we interact with a new website, we may want to interact with it, and to do so, we have to register to create a profile. In this phase, we have to give our data to the company, and they will be stored in their databases.

Problem identification. Let us now try to answer the previous questions to check how the present context is managed:

1. **Interoperability**: we could have two cases. In the first one, we use a technology that enables us to use our existing account on multiple websites, which integrates this solution, for example, "Sign-Up with Google". Here our data is owned by

Google, which shares them (if we grant permission) with third parties, and no one prohibits third parties from keeping our shared data saved. In the second case, we must register each time if the third party does not integrate other "Sign-Up with *" solutions. In both cases, third parties can collect our data (in the first case, Google explicitly knows our interests, but there is a minimum degree of interoperability). Also, in most cases, companies will let us create multiple accounts without verifying our data (one exception to this is the use of KYC).

2. **Privacy**: in some cases, we must show our data with complete transparency: for example, the police stop us on the street and ask for our details. Nevertheless, let us suppose we want to demonstrate something without revealing the details. For example, someone has graduated and wants to demonstrate it without revealing his final grade. We can do this thanks to a cryptography method called *Zero-Knowledge Proof*. However, this has not yet been implemented in most current systems.

3. **Law compliance**: if we consider saving users' data in blockchains, this problem does not exist as we examine centralized systems which do not use them. By the way, of course, there are privacy laws companies must follow (like GDPR).

4. **Security**: our information is stored in databases. With a data breach, considering a centralized system, a malicious actor can access all users' data at once. Sadly, this happens often. So often that someone has made a website where anyone can check if his data has been stolen online at least once (Have I Been Pwned?).

Problem statement. With the above considerations, it is clear that the existing systems work but could be significantly improved. In fact, interoperability enhancement would mean privacy and security penalization. Compromises exist, but if the system is well designed, they can be significantly reduced or at least moved to less dangerous areas. Here, the need for a more secure way to store user data arises. A way that intersects the analyzed points, bringing new power to people and reducing that of companies. This is the Self-Sovereign Identity's principle, which the developed solution will leverage.

Approaches. SSI concept is pretty simple, as opposed to its (in development) implementation. Everyone has different relationships or unique sets of identifying information. This information could include birth date, citizenship, university degrees, or business licenses. In the physical world, these are represented as cards and certificates that the identity holder holds in their wallet or a safe place like a safety deposit box. They are presented when the person needs to prove their identity or something about it. Self-sovereign identity (SSI) brings the same freedom and personal autonomy to the internet in a safe and trustworthy identity management system. SSI means the individual (or organization) manages the elements that make up their identity, and he digitally controls access to those credentials, called Verifiable Credentials (or VCs). They are digital representations of information that can be verified by a third party.

This is achievable by involving three participants:

1. **Holder**: the holder is an individual in the scenario, although it can also be an organization/company. The holder is the entity that holds the credential.

2. **Issuer**: the issuer is the institution, be it a company, certifier body, or governmental organization, that has been awarded a level of trust to provide information (i.e., a public body that issued a passport)

3. **Verifier**: the verifier is the individual, organization, company, or government with whom the holder must prove information's legitimacy and trustworthiness.

The Verifiable Data Registry grants the trust: here are stored schemas and identifiers (linked to the credentials) that the verifiers use to check data validity without the issuer's intervention.

To make a preliminary check of this solution's viability, let us try to answer the previous four questions, considering the new scenario:

1. **Interoperability**: with standards definition, credentials can be presented to verifiers by holders, in the same manner each time. Examples of standards could be credentials schemas (e.g., defining which fields are mandatory) and verification policies (i.e., how the credentials are verified).

2. **Privacy**: as already stated, we can demonstrate something without revealing its details with *Zero-Knowledge Proofs*. This technology has already found applications and implementations in blockchains (e.g., mixers, ZK-rollups, or ZK-games like Dark Forest), so a decentralized system that leverages ZKPs is buildable.

3. **Law compliance**: as will be read later in the paper, this is one of the most challenging points of the full SSI integration with blockchains because of its transparency nature. Everything is registered and immutable, so we must choose what to register and what not. Again, compromises are needed.

4. **Security**: as users hold credentials, as long as they are not saved in centralized servers, significant data breaches (targeting databases) would happen way less often. The user is responsible for his information security, and secure communication protocols will enhance it.

After quickly drafting these reflections, it can be said that SSI principles fit our problem requests, so a solution that aims to solve them can be tried to be developed.
These analyzed points are well discussed in an article by Christopher Allen called "The path to Self-Sovereign Identity", where he defines the "Ten Principles of Self-Sovereign Identity".

1.2 Basic use cases

After addressing the problem and trying to provide answers to the initial questions, it is possible to start thinking about the first use cases. Obviously, the minimum requirement is credentials involvement: each time a user has to demonstrate some information, SSI could theoretically be leveraged.

In the first part of the internship, the focus has been (after the SSI primitives study) on use cases. They can be grouped into these two macro-categories:

Academics Here can be included all the uses about, for instance, university. A lot of them can be thought about and analyzed, but the most interesting which have been examined are these:

1. **Exams and Diplomas emission**: students can present their credentials (badge) to the university to register for exams. Each exam result would be another credential, and in order to access the diploma, he should present all the credentials related to passed exams (possibly wrapping them in a "presentation"). The final diploma would be another verifiable credential emitted by the university.

2. **Scholarships requests**: students can demonstrate they are eligible for facilitations by presenting their credentials to the university. This way, information pieces are easily checkable and verifiable, and procedures would be faster and less susceptible to errors. A "permit" credentials could be emitted, which would grant the student access to facilitations (e.g., canteen, money, discounts...)

3. **Discounts and university canteen**: it is evident that comfortable functions come into effect by processing the previous use case. Instead of creating accounts with university e-mail, students should present their credentials to services that offer facilitations, bringing interoperability and trust to each part.

Institutionals Here can be placed all use cases involving the participation of national, European (or other unions), or global entities. Noteworthy examples could be:

1. **National ID**: this new system would replace physical identity cards with verifiable credentials. The municipality would issue them to the citizens, and the latter would use them to access all the national services or other services which request IDs.

2. **European SSI system**: this use case is currently in development and is called EBSI (European Blockchain Services Infrastructure). The final aim is to introduce the use of verifiable credentials in Europe and introduce new types of services to European citizens or improve the current ones.

These are just some examples of the possible use cases that can be developed, considering the model SSI offers. With the final product built during the internship (obviously, this would need additional integrations but gives a solid base), those listed are all viable scenarios.

1.3 Internship description

The previous sub-chapters clearly defined the problems and use-cases the product aims to involve and realize. This one will describe the structure of the internship, the company where it took place, and how the work was done.

1.3.1 The company

Officially, the company where the internship took place is called Athesys (from the Latin version of "Adige", i.e., "Athesis"), but actually, the job was about their startup, called **Monokee**.

Athesys is a XaaS (Anything as a Service) integrator founded in 2010; they provide services such as database management, business intelligence, software development, security, and cloud.
In 2012 they began thinking about IAM (Identity and Access Management) solutions, delivering a product that, among many things, provides a Single Sign-On functionality across different domains.

Monokee is born in 2017, and it is an innovative product-oriented startup that serves as an IAM for centralized and decentralized digital identities. In fact, its solution is hybrid: to the classic method (which involves using databases to store information), it intends to add SSI techniques, and this is where the internship comes in.

1.3.2 Internship objectives and planning

It is dutiful to specify that, at least initially, the objectives were not crystal clear. The overall concept of the internship itself was well defined, but the path to developing the whole solution was not.
The main objective was to develop **a software that enables the user to interact with verifiable credentials**. This type of software is named *Agent*, and users are intended as holders, issuers, and verifiers.
Before the internship, the Monokee team searched for some existing solutions (unfortunately, not numerous) and found an SSI Kit developed by walt.id, a European company focused on SSI.

So, in the beginning, the steps to follow were:

1. Deep dive into SSI technology and its primitives;
2. Analysis of the problem and understanding of what is needed and how to use it for the final product;
3. Study of SSI Kit, provided by walt.id;
4. If it fits the needs, leverage it to develop the Agent (if not, develop a similar software);
5. If possible, integration into a web application Proof of Concept with blockchain's smart contracts.

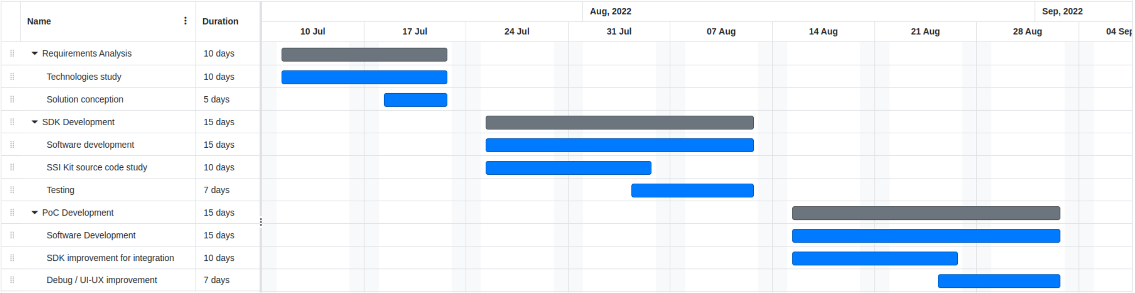
The path to pursue became more apparent during the first two steps (technology study and problem analysis), so the final internship structure became this:

1. **Requirements analysis**: in this phase, SSI has been well studied and comprehended to understand the next steps. It has been divided into:
   (a) **Technologies study**: understanding of the existing standards;
   (b) **Solution conception**: definition of the following steps;
2. **SDK Development**: development of the library that serves as an abstraction of the existent SSI Kit, allows it to be used on a web application. Divided into:
   (a) **Software development**: code development of main entities;

   (b) **SSI Kit source code study**: needed mostly because of documentation lack;
   (c) **Testing**: unit testing of the library main components;

3. **PoC Development**: final part of the internship, where has been developed a web application that merges the SSI Kit SDK with smart contracts with SSI features. Separated into:
   (a) **Software development**: development of the web application (back-end and front-end);
   (b) **SDK improvement for integration**: improvement of the SDK to better fit the web application needs;
   (c) **Debug/UI-UX improvement**: final arrangements of the proof of concept.

In the following Gantt chart are outlined the timings of each step:



Name	Duration
Requirements Analysis	10 days
Technologies study	10 days
Solution conception	5 days
SDK Development	15 days
Software development	15 days
SSI Kit source code study	10 days
Testing	7 days
PoC Development	15 days
Software Development	15 days
SDK improvement for integration	10 days
Debug / UI-UX improvement	7 days

Figure 1.1: Internship structure

Chapter 2

State of the art and technology background

This chapter presents the pre-concepts needed to comprehend this paper's content fully. As is understandable from the introduction, they are about Self-Sovereign Identity and blockchains. In addition, state of the art will be analyzed to see what has already been done and what can be improved.

2.1 Technology concepts

This section will explain in detail SSI and blockchain technologies.

2.1.1 Self-Sovereign Identity concepts

Here can be read a brief reprise of what has already been saying about Self-Sovereign Identity and a description of its main primitives: VCs, VPs, and DIDs.

Self-Sovereign Identity

Self-Sovereign Identity is an approach to digital identity that gives individuals control over their data. SSI addresses the difficulty of establishing trust in interaction and allows people to interact in the digital world with the same freedom and ability to trust as they have in the offline world.

To be trusted, a party in an interaction will present credentials to other parties, and those parties can verify that the credentials come from an **issuer** they trust. This way, the **verifier**'s trust in the issuer is transferred to the credential **holder** (or **prover**). This basic structure of SSI with three participants is sometimes called the "triangle of trust.", simply because you need an element of trust among these entities for them to work together.

While this does not mean that there is a legal partnership or understanding between the entities involved, it does mean that each of the entities is willing to examine the credibility of the other, and this implicit trust is what constitutes this term.

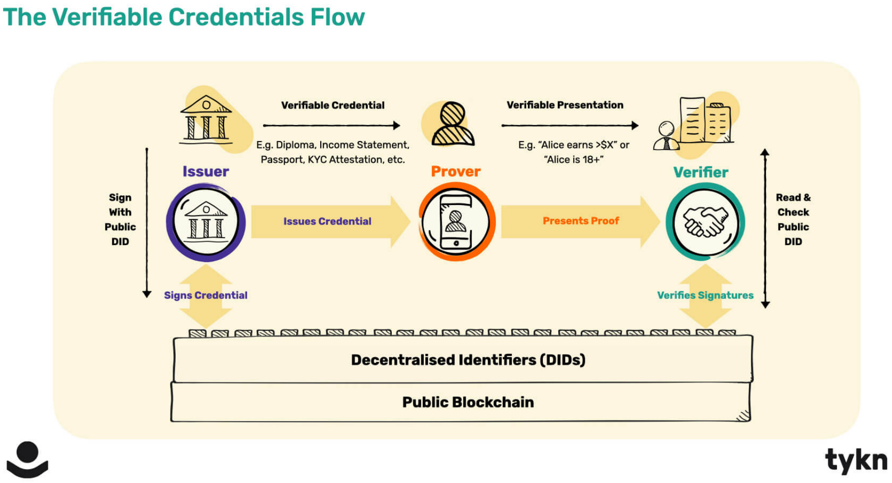
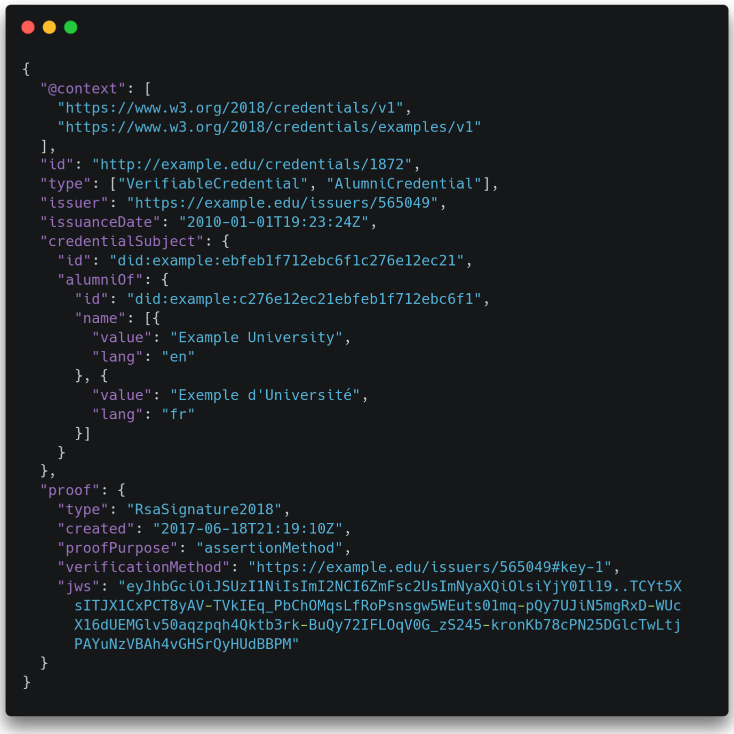


Figure 2.1: The triangle of trust: Prover, Issuer, and Verifier (by Tykn)

Verifiable Credential (VC)

A verifiable credential can represent all of the same information that a physical credential represents. The addition of technologies, such as digital signatures, makes verifiable credentials more tamper-evident and more trustworthy than their physical counterparts.



```
{
  "@context": [
    "https://www.w3.org/2018/credentials/v1",
    "https://www.w3.org/2018/credentials/examples/v1"
  ],
  "id": "http://example.edu/credentials/1872",
  "type": ["VerifiableCredential", "AlumniCredential"],
  "issuer": "https://example.edu/issuers/565049",
  "issuanceDate": "2010-01-01T19:23:24Z",
  "credentialSubject": {
    "id": "did:example:ebfeb1f712ebc6f1c276e12ec21",
    "alumniOf": {
      "id": "did:example:c276e12ec21ebfeb1f712ebc6f1",
      "name": [{
        "value": "Example University",
        "lang": "en"
      }, {
        "value": "Exemple d'Université",
        "lang": "fr"
      }]
    }
  },
  "proof": {
    "type": "RsaSignature2018",
    "created": "2017-06-18T21:19:10Z",
    "proofPurpose": "assertionMethod",
    "verificationMethod": "https://example.edu/issuers/565049#key-1",
    "jws": "eyJhbGciOiJSUzI1NiIsImI2NCI6ZmFsc2UsImNyaXQiOlsiYjY0Il19..TCYt5X
      sITJX1CxPCT8yAV-TVkIEq_PbChOMqsLfRoPsnsgw5WEuts01mq-pQy7UJiN5mgRxD-WUc
      X16dUEMGlv50aqzpqh4Qktb3rk-BuQy72IFLOqV0G_zS245-kronKb78cPN25DGlcTwLtj
      PAYuNzVBAh4vGHSrQyHUdBBPM"
  }
}
```

Figure 2.2: Example of verifiable credential (VC)

Holders of verifiable credentials can generate verifiable presentations and then share these verifiable presentations with verifiers to prove they possess verifiable credentials with certain characteristics.
Both verifiable credentials and verifiable presentations can be transmitted rapidly, making them more convenient than their physical counterparts when trying to establish trust at a distance. The three main components of a VC are:

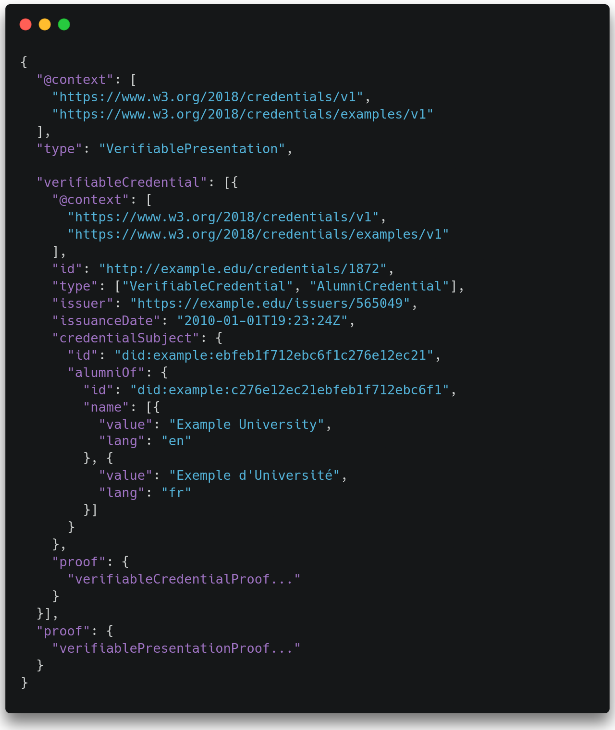
1. **Metadata**: cryptographically signed by the issuer. It describes the credential properties, such as the issuer, the expiry date and time, a public key to use for verification purposes, the revocation mechanism, and other information;

2. **Claims**: a statement made about a subject. Example: "Janice's date of birth is 01/01/1990."

3. **Proofs**: a proof is data about the identity holder that allows others to verify the source of the data (i.e., the issuer), check that the data belongs to (only) the holder, that the data has not been tampered with, and finally, that the issuer has not revoked the data.

Verifiable Presentation (VP)

A verifiable presentation expresses data from one or more verifiable credentials and is packaged in such a way that the authorship of the data is verifiable. If verifiable credentials are presented directly, they become verifiable presentations. Data formats derived from verifiable credentials that are cryptographically verifiable but do not themselves contain verifiable credentials might also be verifiable presentations.



```
{
  "@context": [
    "https://www.w3.org/2018/credentials/v1",
    "https://www.w3.org/2018/credentials/examples/v1"
  ],
  "type": "VerifiablePresentation",

  "verifiableCredential": [{
    "@context": [
      "https://www.w3.org/2018/credentials/v1",
      "https://www.w3.org/2018/credentials/examples/v1"
    ],
    "id": "http://example.edu/credentials/1872",
    "type": ["VerifiableCredential", "AlumniCredential"],
    "issuer": "https://example.edu/issuers/565049",
    "issuanceDate": "2010-01-01T19:23:24Z",
    "credentialSubject": {
      "id": "did:example:ebfeb1f712ebc6f1c276e12ec21",
      "alumniOf": {
        "id": "did:example:c276e12ec21ebfeb1f712ebc6f1",
        "name": [{
          "value": "Example University",
          "lang": "en"
        }, {
          "value": "Exemple d'Université",
          "lang": "fr"
        }]
      }
    },
    "proof": {
      "verifiableCredentialProof..."
    }
  }],
  "proof": {
    "verifiablePresentationProof..."
  }
}
```

Figure 2.3: Example of verifiable presentation (VP)

The data in a presentation is often about the same subject but might have been issued by multiple issuers. The aggregation of this information typically expresses an aspect of a person, organization, or entity.

Decentralized Identifier (DID)

JSON, JWS and JWT

2.1.2 Blockchain concepts

Blockchain

Permissionless and permissioned blockchains

Bitcoin

Ethereum

Hyperledger

Hyperledger Besu

Hyperledger Fabric

2.1.3 Libraries and Stack involved

EBSI

walt.id SSI Kit

2.2 State of the art

Chapter 3

Solution

3.1 Solution proposal

3.2 Solution development

3.2.1 Technologies and Tools

3.2.2 SSI Kit SDK development

3.2.3 Smart Contract integrations - Web App Proof of Concept

3.3 Discussion

3.3.1 Achievements

3.3.2 Acquired knowledge

3.3.3 Future developments

3.3.4 Personal evaluation

Conclusion

Bibliography

Bibliographical references

James P. Womack, Daniel T. Jones. *Lean Thinking, Second Editon.* Simon & Schuster, Inc., 2010.

Websites consulted

Manifesto Agile. URL: http://agilemanifesto.org/iso/it/.