NGI ONTO CHAIN

Blockchain for the Next Generation Internet

D3.1 STATE OF THE ART ANALYSIS REPORT

08/11/2021

Grant Agreement No.: 957338
Call: H2020-ICT-2020Call: H2020-ICT-2020
Topic: ICT-Topic: ICT-54
Type of action: RIA2020
Type of action: RIA

D3.1 STATE OF THE ART ANALYSIS REPORT

Work package	WP3
Task	T3.1
Due date	31/08/2021
Submission date	08/11/2021
Deliverable lead	IS
Version	0.8
Authors	Thanasis Papaioannou (AUEB), Klevis Shkembi (UL), Petar Kochovski (UL), Alberto Ciaramella (IS), Vlado Stankovski (UL)
Reviewers	Marco Ciaramella (IS), Caroline Barelle (ED)
Abstract	This deliverable summarizes the status of the art which has been taken into account in the definition of the ONTOCHAIN architecture, for blockchain, semantics and both. This status of the art includes information from scientific and business sources, from standards, from patents and from other research projects.
Keywords	Blockchain, Semantic Web, Ontologies, Knowledge Management

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Document Revision History

Version	Date	Description of change	List of contributors
0.1	27/07/2021	Table of Contents	Alberto Ciaramella
0.2	15/08/2021	Section 1 (draft) + Section 7 (in part)	Alberto Ciaramella
0.3	23/08/2021	Section 2 (in part), Section 3 (in part)	Klevis Shkembi, Alberto Ciaramella
0.4	25/08/2021	Section 1, Section 2, Section 3, Section 4, Section 5, Section 6	Thanasis Papaioannou, Vlado Stankovski, Petar Kochovski, Klevis Skempi
0.5	30/9/2021	Section 8, Section 10	Thanasis Papaioannou
0.6	29/10/2021	Executive Summary, editorial fixes	Thanasis Papaioannou
0.7	4/11/2021	Internal Review	Marco Ciaramella, Caroline Barelle
0.8	8/11/2021	Applied reviewer comments	Thanasis Papaioannou

Dissemination Level

Nature of the deliverable:		PU
PU	Public, fully open, e.g., web	*
CL	Classified, information as referred to in Commission Decision 2001/844/EC	*
CO	Confidential to ONTOCHAIN project and Commission Services	*

DISCLAIMER

The information, documentation and figures available in this deliverable are written by the "Trusted, traceable and transparent ontological knowledge on blockchain — ONTOCHAIN" project's consortium under EC grant agreement 957338, and do not necessarily reflect the views of the

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

The ONTOCHAIN project is funded by the European Union's Horizon 2020 Research and Innovation programme under grant agreement no. 957338.

COPYRIGHT NOTICE

The ONTOCHAIN Consortium is the following:

Participant number	Participant organization name	Short name	Country
1	EUROPEAN DYNAMICS LUXEMBOURG SA	ED	LU
2	UNIVERZA V LJUBLJANI	UL	SI
3	IEXEC BLOCKCHAIN TECH	IEXEC	FR
4	INTELLISEMANTIC SRL	IS	IT
5	ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS - RESEARCH CENTER	AUEB	EL
6	ELLINOGERMANIKO EMPORIKO & VIOMICHANIKO EPIMELITIRIO	GHCCI	EL
7	F6S NETWORK LIMITED	F6S	IE

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

EXECUTIVE SUMMARY

This report aims to overview the state of the art in the full spectrum of technologies that will be studied and employed for the development of the ONTOCHAIN platform. The ONTOCHAIN project aims to combine blockchain with data semantics, so we overview the main concepts in the distributed ledger technologies and in the semantic web. Based on the high-level requirements of the innovative applications to be supported by the ONTOCHAIN platform, we derive a high-level architectural view for the platform. The state of the art related to each functionality of the different building blocks of this architecture is studied and analyzed. Moreover, the verticals and their respective indicative applications that benefit from blockchain are also reviewed to select use cases for demonstrating the added value of the ONTOCHAIN platform. Standardization activities in the blockchain domain have also been discussed and some challenges for the realization of the ONTOCHAIN platform have been identified. Finally, based on the feedback obtained by the members of our Advisory Board, which is discussed in detail, we verify that the ONTOCHAIN vision and its associated technologies have been adequately considered in the ONTOCHAIN project so far.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

TABLE OF CONTENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

LIST OF FIGURES

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

LIST OF TABLES

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

ABBREVIATIONS

CEN	Comité Européen de Normalisation
CENELEC	Comité Européen de Normalisation Électrotechnique
DAO	Decentralized Autonomous Organization
DLTs	Distributed Ledger Technologies
ETSI	European Telecommunications Standards Institute)
HIMSS	Healthcare Information and Management Systems Society
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
ISO	International Organization of Standardization
ITU	International Telecommunication Union
SC	Smart Contract
SDO	Standard developing organizations
SLA	Service Level Agreement
SSI	Self sovereign Identity
SSO	Single sign on
VC	Verifiable Credentials
W3C	World Wide Web Consortium

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

1. INTRODUCTION

1.1 THE ONTOCHAIN TECHNOLOGY

The Next Generation Internet initiative aims at making the Internet more human centered with openness, inclusivity, transparency, privacy, cooperation, and protection of data. The ONTOCHAIN project contributes to this vision through the development of a software ecosystem for trustworthy knowledge and information handling and trustworthy content exchange with a specific focus on Blockchains, Smart Contracts, Decentralized Oracles, Semantic Web, Ontologies and related knowledge-management technologies.

ONTOCHAIN envisions to create a multi-layer and modular blockchain framework, to enable the implementation of a number of different next-generation real-world solutions, such as trustworthy web and social media, trustworthy crowdsensing, trustworthy service orchestration, unsupervised, decentralized online social networks, etc. and to empower practitioners to address the various challenges of the Internet (e.g., centralization of power and knowledge, unknown provenance of information, anonymous and unreliable identifiers, personal data exploitation, AI biases, data censorship, fraud, etc.) through the use of multiple ledger and semantic technologies.

Our use-cases will rely on successful Semantic Web approaches such as Linked Data, OWL Lite, OWL DL and other approaches and formats that will deliver a trustworthy, privacy-preserving, secure, transparent, democratic and traceable approach to manage access and operations over ontologies, metadata, data, knowledge and information in the ecosystem. Of primary importance is to well understand these technologies and their key attributes to demonstrate their relevance for promoting user trust and helping create new ethical and virtuous business opportunities benefiting citizens, companies, and public services.

The ONTOCHAIN framework design is delivered as part of the main ONTOCHAIN consortium activities. Moreover, the ONTOCHAIN consortium defines specific technical tasks for the detailed design and the implementation of the ONTOCHAIN framework and outsources these tasks to third-party applicants in three open calls for subprojects.

Selected sub projects as part of the three ONTOCHAIN open calls in collaboration with ONTOCHAIN consortium members and along with guidance and mentoring deliver the technical components (algorithms and modules) of the ONTOCHAIN framework. More specifically, calls 1 and 2 will be oriented to develop the technical infrastructure, whilst call 3 will deliver new applications.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

1.2 PURPOSE AND STRUCTURE OF THIS DOCUMENT

This report summarizes the state of the art relevant for the ONTOCHAIN project extracted from the literature and the industry with the following objectives:

1) To identify the main technological challenges that must be faced in ONTOCHAIN towards its goals. Based on a preliminary design of the ONTOCHAIN architecture to identify its key functional components and their associated technological domains to be reviewed.
2) To overview main distributed ledger technologies and the semantic web technologies related to blockchain.
3) To overview the state of the art in the various areas that comprise the functionalities to be provided by the ONTOCHAIN framework.
4) To overview the various application domains (i.e., verticals) and specific application examples currently employed in the blockchain and identify potential for enhanced ONTOCHAIN applications. This, however, will be fully addressed in our subsequent deliverable D3.2.
5) To overview the various standardization activities in the blockchain domain and ONTOCHAIN participation in those.
6) To state the main technological challenges towards the realization of ONTOCHAIN.
7) To describe the feedback collected by the Advisory Board in the activities of the first year of ONTOCHAIN.

The remainder of this document is organized as follows:

- **Section 2** discusses the challenges of the current Internet that ONTOCHAIN aims to tackle and describes in high-level the ONTOCHAIN framework to tackle them.
- **Section 3** describes the fundamentals of blockchain and the main blockchain platforms.
- **Section 4** overviews semantic web technologies relevant to blockchain.
- **Section 5** overviews state of the art related to the functional blocks of the preliminary ONTOCHAIN framework.
- **Section 6** discusses the potential use cases of ONTOCHAIN in the various verticals with respect to the current practice.
- **Section 7** overviews the standardization efforts in the blockchain domain.
- **Section 8** identifies the main technical challenges towards the realization of the ONTOCHAIN framework.
- **Section 9** describes the feedback obtained by the Advisory Board of ONTOCHAIN for the project progress in its first year.
- Finally, **Section 10** concludes this report.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

2. THE ONTOCHAIN TECHNOLOGICAL FRAMEWORK

2.1 THE INTERNET CHALLENGES AND ONTOCHAIN

Today, the Internet is involved in all aspects of our lives. A significant portion of the Internet is the World Wide Web, and its services, which are provided by various giant, large and small legal entities and individuals.

With the number of services available constantly on the rise, we are witnesses to an ever-increasing information overload. In addition, poor content aggregation mechanisms and stovepipe systems are making effective collaboration and smart decision making an even bigger challenge.

Notwithstanding the ability of advanced technologies to distinguish factual from non-factual data, existing large or small WWW services are used today with the purpose of spreading misleading information that usually serve a certain purpose: to damage one's reputation, win an election, make people buy products and services. With the confluence of the WWW with the Internet of Things, the ubiquitous Artificial Intelligence, the existence of Cloud, Fog and Edge computing platforms and similar, it becomes apparent that the existing problems of misuse of information can soon achieve even more dangerous levels of potential manipulation of the people that must be prevented.

As a response to these challenges a new vision has arisen. A vision where Internet (WWW, social networks, social media and IoT, etc.) data are understood by the machines and made accessible to an array of semantic technologies, therefore allowing the machines to do more effective and value adding work when responding to service requests.

Technically, this is achieved by using ontologies, that is, "formal, explicit specification of shared conceptualizations". Ontologies make it possible to intertwine the data and information into a Web of Knowledge. Several successful companies have built on the Semantic Web ideas in the past decades and have had enormous success, with the most popular applications being in the form of knowledge graphs such as Google Knowledge Graph or IBM Socrates.

However, the Semantic Web does not execute uniformly for all. In such a system actors can sometimes make completely opposed assertions, such as "that apple is red" and "that (same) apple is yellow". This concept becomes especially important in crowdsensing which allows anyone to contribute the data acquired by their own connected objects in order to build collaborative knowledge. What is currently necessary, is to be able to establish the truth from several assertions.

With the emergence of the Internet of Things (IoT), the new wave of Artificial Intelligence (AI), Orchestration and novel Cloud Continuum approaches (Edge, Fog, Data Center), we now have the potential to reach

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

a new level of decentralization, but also of cooperation between various cyber-physical systems based on the Semantic Web principles. Blockchain technologies with their main properties of decentralisation, traceability and transparency fit perfectly to this agenda, and may contribute to achieving trusted operations of such smart applications and systems (Kochovski, 2019).

The hypothesis of this work is that with these intrinsic properties of blockchain, it is possible to establish a common, shared ledger for the management of shared ontological concepts including instances of such concepts. An important aspect of ONTOCHAIN is the ability to interlink off-chain data, information and (AI) services with on-chain information in a way that reduces the need for costly on-chain operations and provides significant new properties, such as traceability, privacy, mechanisms for democracy and other.

Membership of different entities (e.g., specific objects, persons), in specific ontological concepts can be established, for example, by means of independent evaluation of various stakeholders with the use of AI methods. These entities can be anonymous, but at the same time, they are able to be linked to real-word identities, when law demands it. Not only ontological concepts may be well-agreed among the participants, but also, they can be directly ''executable'' through the employment of various semantic reasoners, operating directly on blockchain, potentially also employing trustworthy off-chain real-world data (e.g., IoT) with the use of Smart Oracles and Decentralized Oracles that establish facts by using democratic, decentralized means.

A multi-layer approach to reach the envisioned ONTOCHAIN framework and to serve the defined use-cases and applications is followed as described in Figure 1. This framework will enable the implementation of several innovative different next-generation real-world solutions, such as trustworthy web and social media, trustworthy crowdsensing, trustworthy service orchestration, unsupervised/decentralized online social networks, etc.

Eventually, we predict that the diversity, the complexity, and the specialization of different real-world ONTOCHAIN applications will lead practitioners to use multiple ledger technologies for implementing different solutions. This will enable higher performance and scalability, while enabling different business logic, access methods and governance models that require specific chains.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

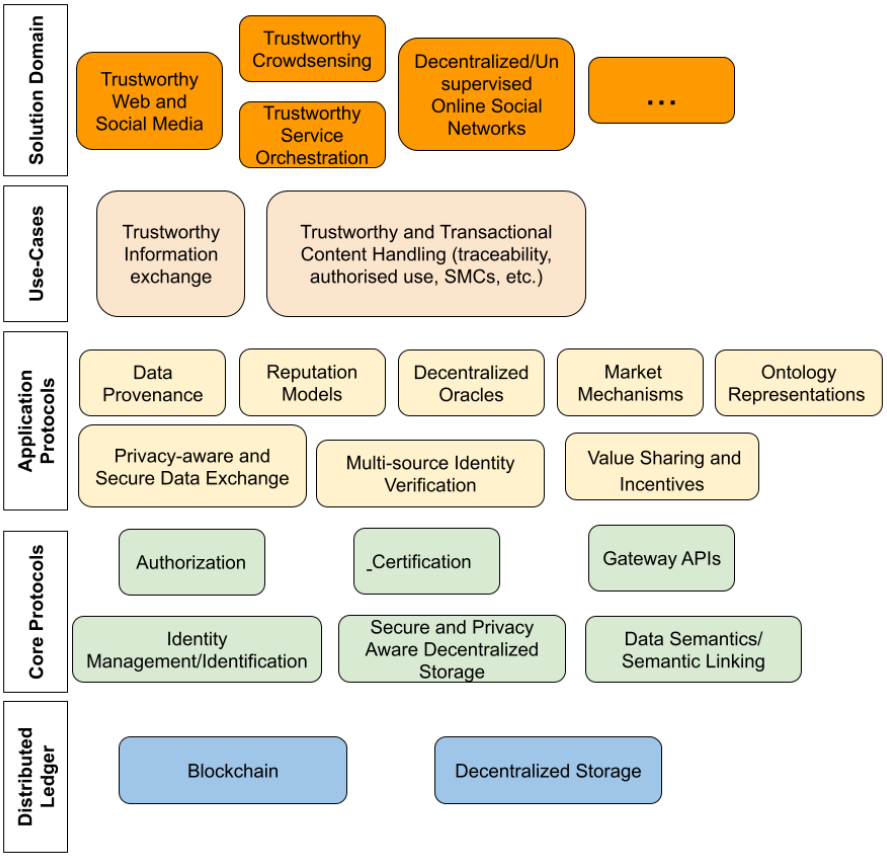

FIGURE 1: THE ONTOCHAIN LAYERED TECHNOLOGICAL FRAMEWORK

ONTOCHAIN use-cases will be built upon the different protocols shown in Figure 1. It is important to note that most of the components of the proposed architecture do not exist into any of the blockchain platforms mentioned in Section 3. ONTOCHAIN Application and Core protocols will implement the interactions between different blockchain frameworks, while hiding them from the use-cases to support effortless inter-service process cooperation. Moreover, data stored at different chains, may be linked together. This linkage will be stored in new ONTOCHAIN chains. Data stored at the chains of ONTOCHAIN is referred to on-chain data, as opposed to external data that is stored outside the ONTOCHAIN chains, which is referred to as off-chain data.

For enabling scalability, openness and high performance, we employ a modular approach. Each of the modules and functionality of each layer is built upon functionality offered by the lower layers. The functionality of the modules at each layer is described in a top-down manner below, along with the dependencies among them. At the Solution Domain layer lie different next-generation application solutions, such as trustworthy web and social media, trustworthy crowd-sensing, trustworthy service orchestration, decentralized online social networks, which tack today's

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Internet problems that can be built upon the use cases Trustworthy Information Exchange and Trustworthy and Transactional Content Handling. Each of the use cases is built upon combined functionality from the Application Protocols layer, such as Data Provenance, Reputation Models, Decentralized Oracles, etc. The modules at the Application Protocols layer themselves are built upon core blockchain-based services at the Core Protocols layer, such as Smart Contracts, Identity Management, Secure and Privacy-Aware Decentralized Storage, Certification, Authorization and Data Semantics. The Core Protocols modules employ basic Distributed-ledger functionality, i.e., Blockchain, Digital Currency and Distributed Storage, which lie on combined proprietary, corporate and public resources. The functionality of the modules at each layer is described in a top-down manner below, along with the dependencies among them.

2.1.1 Use-Case Layer

Trustworthy Information Exchange: This use case encapsulates the tools and libraries for the secure exchange of trustworthy data among trustworthy parties. It employs and combines data provenance mechanisms, decentralized oracles and user trustworthiness to assess trustworthiness of information. Decentralized reputation models are employed to assess the trustworthiness of data sources and that of the data itself, while the secure data exchange mechanisms are employed to transfer the data securely among transacted parties through cryptographic mechanisms.

Trustworthy and Transactional Content Handling: This use case enables trustworthy data handling by means of any combination of the following: authorized access/handling of the data, data credibility assessment, implementation of copyrights, secure and privacy aware querying of the data (e.g., by means of secure multiparty computation and data sanitization approaches). It addresses softer requirements for content handling where the decision on how to operate in certain situations may differ on a case-to-case basis. Moreover, data transactions involve some trading value, which is going to be assessed by means of underlying market mechanisms, while the generated economic benefit should be shared among different contributors by means of economic mechanisms in a fair and incentive-compatible manner.

This use case also deals with the secure transfer of any financial assets among involved parties in a data transaction. Regulatory alignment of data transactions, as a part of Trustworthy and Transactional Content Handling, will define and develop tools and mechanisms that would allow regulatory, judiciary and law enforcement agencies to introspect and otherwise influence data transactions in strictly defined circumstances envisioned by legislature.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

2.1.2 Application Protocols Layer

Data Provenance:

This module will provide graphical and programming interfaces for querying and presenting provenance information from ONTOCHAN about on-chain and off-chain data. Provenance information will include the complete trail of transactions that resulted in a record, including links to components that were run (e.g., signature of AI models when available), to the input data that was processed and to the contributors who ran the programs or provided original information.

Reputation Models:

This module will provide the functionality of building different decentralized reputation models over the blockchain, so that reputation feedback is genuine, credible, and anonymous. The basic building blocks of a reputation system are an approach for casting assessments/votes for a particular subject (person/data/fact), an approach for recording the history of votes per subject and an approach for summarizing votes into a single reputation metric per subject. This module is built upon Identity Verification mechanisms.

Decentralized Oracles:

This module will facilitate Smart Contracts to operate with off-chain data, although by design, Smart Contracts can only read and write data that is stored on their blockchain. By design, Smart Contracts can only read and write data that is stored on their blockchain. This property is fundamental to blockchains: if Smart Contracts could read any data, their execution would not be deterministic, and no consensus on their transactions could ever be reached.

To avoid centralization, some approaches (e.g., Substrate, ChainLink) apply multiple instances to look at a data source, and then run a consensus algorithm on-chain to validate the result. This, however, only displaces the point of centralization from the Oracle to the data source.

While the idea of Decentralized Oracles is simple, its implementation is not trivial: every use-case requires different data sources, and the consensus algorithm based on multiple data types can become complex.

Market Mechanisms "as a Service":

This module provides the basic support mechanisms for enabling data/service transaction, and thus enables market mechanisms. For example, this module will support trading of physical assets (e.g., tokenization) and price determination (e.g., auctions, negotiation protocols, etc.), billing, customer support, inventory management services and more. It also provides functionality for enabling the

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

sharing economy, such as value chaining, value/cost sharing and DeFi support.

Secure Data Exchange:

This module comprises the functionality of exchanging data among distributed parties, while verifying the ownership of the data and access rights, authenticity of transacted parties, the integrity of the data exchanged and the confidentiality of the data through blockchain underlying mechanisms. Most often, off-chain data will be exchanged in data transactions, while on-chain data will store public cryptographic keys and access control lists based on which elevated data access to different portions of data is authorized for specific transacted parties.

Ontology Representation:

This module seeks to define new ways for implementing ontologies with the use of blockchain. Semantic agreements can be commonly agreed based blockchain-based consensus, similarly to the establishment of axiomatic statements. Moreover, new ontologies will be defined for smart contracts and decentralized services to enable service searchability and matching with service requests. This module will also include any reasoning approaches, tools and methods that can help deduce new knowledge arriving from a sensing IoT empowered environment.

Multi-source Identity Verification:

This module seeks to register and verify individual digital identities of physical objects via newly designed ONTOCHAIN services. The potential of attributes assertion on ONTOCHAIN is tectonic in nature. Attributes represent a fundamental part of any ontological concept.

For instance, various AI methods could be introduced to operate on sensing data (IoT based, sensors, cameras and similar) to assert whether an individual belongs to a specific ontological concept.

Value Sharing and Incentives:

ONTOCHAIN ecosystem is to be, by nature, a public good built upon the resources and efforts of a great number of people. Proper incentive mechanisms for rewarding the people involved, according to their contribution, should be in place. Such mechanisms could include: i) the generation of a certain number of cryptocurrencies for block mining and execution of smart contracts, ii) contribution assessment. These are facilitated through an appropriate accounting system for measuring resource consumption for blockchain tasks.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

2.1.3 Core Protocols

Smart Contracts:

These are computer programs that can execute transactions directly on blockchains when some predefined conditions are met. Some Smart Contracts allow the inclusion of oracles to make decisions on transactions based on real world information (i.e., outside of the blockchain).

Certification:

This module refers to the confirmation of certain characteristics of an object, person, or organization. For example, a government may decide to offer certificates to cloud providers that have verified GDPR-compliant handling of private citizens' data. In such case, certificates can be issued on-chain (i.e., implemented within Smart Contracts), and can be used as conditions for performing specific transactions, for example, using AI methods to analyze private data. Hence, specific conditions can be implemented within a Smart Contract to govern the GDPR-handling of private citizens' data only on certified cloud providers.

Secure/Privacy Aware Storage:

This module encapsulates solutions already existing on blockchain. Together with decentralisation they help reduce the risk of one-party having access to all private data. Moreover, various partitioning, fragmentation and redundancy methods will be used (e.g., StorJ).

Identity Management:

This module deals with technologies and solutions to address parts of the digital identity puzzle. There are two conflicting requirements that drive this development: i) ability to identify oneself in specific interactions (e.g., withdrawing money in a bank), ii) preservation of one's privacy (e.g., healthcare data, online buyer's habits). This is a feasible endeavor. However, it is necessary to invest more in technologies like ONTOCHAIN to make it happen.

Optimization:

This will provide new semantics-related solutions and will seek to minimize the necessary amount of both on-chain and off-chain transactions, to reduce the operational cost and improve its overall efficiency, including energy-efficiency. Because the ontology and semantic reasoning mechanisms will be built on top of a blockchain, all data it contains will be irreversibly stored by default. The critical issue to address here are the new algorithms that would achieve the same or similar level of trustworthiness, provenance, and other effects, while reducing the number of On-ONTOCHAIN transactions.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Gateways/Bridges:

This module will support connections between the ONTOCHAIN blockchain and the outside world, including other blockchains in the form of Smart Contracts, as well as several higher-level wrappers for commonly used languages (e.g., JavaScript, Java and Python). Part of its duty will be to help the engineers in the upper layers to make tradeoffs about how much data is stored On-Chain, by supporting pointers to Off-Chain decentralized storage, such as IPFS. The module will provide several low-level Application Programming Interfaces (APIs) in the form of Smart Contracts, as well as several higher-level wrappers for commonly used languages (e.g., JavaScript, Java and Python). The interfaces will be generic and extensible to allow connections with different ledger technologies in the future, while only external Ethereum-based chains will be supported during the course of ONTOCHAIN.

Our prototype will be implemented using the Ethereum software stack, because of its important community of adopters and developers.

Data Semantics:

Since ontology engineering is a complex work that usually takes many years to complete and test, this module intends to stimulate reuse of this body of generated knowledge in order to foster the use of various schemata and ontologies when describing the semantics of data.

Ontologies are core building block of the Semantic Web[1]. The W3C consortium provides mechanisms for their standardization to foster their use in applications world-wide, with the potential to build various artificial agents that can cross-link the information and perform advanced queries via SPARQL. Supporting these standards in the blockchain and providing data semantic annotation, semantics extraction, linking, inference, alignment, and reasoning on top of blockchains will significantly boost the business viability of future applications involving knowledge management.

Authorisation:

Blockchain has stimulated the idea of self-sovereign digital identity and few commercial services have already emerged[2]. Various Role-Based Access Control (RBAC) systems have also existed for decades. With this module, one could easily see systems where a patient is self-identified on blockchain, while a medical doctor gains access to the medical records based on her/his role (e.g., surgeon, general practitioner).

[1] https://www.w3.org
[2] https://www.ibm.com/blogs/blockchain/category/trusted-identity/self-sovereign-identity/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

2.1.4 Distributed ledger

Blockchain Consensus Engine:

Consensus making mechanisms are at the core of any blockchain. ONTOCHAIN will be designed to be scalable, open, cost and energy-efficient, and when possible, even as a much-improved new consensus engine. A consensus engine that determines consensus in blockchain writing in a scalable and irrefutable way is on the research agenda of many, and ONTOCHAIN poses significant new requirements for such design.

Regarding openness, ONTOCHAIN does not aim for a silo blockchain ecosystem, but for an open distributed ledger that in principle can be combined with different blockchain environments. Therefore, consensus-making mechanisms should not be bound to any specific API requirements for the distributed ledger.

Digital Currency:

This are praised for allowing cheap and fast money transfer to the 1.7 billion people who are excluded from the banking system around the world, or as a stable alternative to devalued fiat currencies. One very interesting aspect for the Next Generation Internet is the possibility of programming complex self-executing transactions in Smart Contracts. Integrated with ONTOCHAIN's provenance and reputation mechanisms, a CryptoToken will guarantee a fair compensation to every contributor who participates in the ecosystem.

Decentralised Storage:

Various decentralised repositories, such as Peer-to-Peer and Content Distribution Networks have existed for decades. With the emergence of blockchain, we have witnessed a new wave of participatory storage repositories that can help address the security and privacy needs and may help store practically any kind of data (e.g., StorJ). Soon, one could imagine new storage services that can help store private data in encrypted and decentralised way, that can help manage data replicas for reliability and Quality of Service, while balancing the trade-offs with the storage costs.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

3. DISTRIBUTED LEDGER TECHNOLOGIES

3.1 BLOCKCHAIN BASICS

Satoshi Nakamoto defined it in 2008 as highly disruptive technology based on the use and benefits of Bitcoins (Nakamoto 2008). It is a series of blocks linked by the previous block's hash value and so on. Block headers, timestamps, block indexes, block hashes, preceding block hashes, Merkle root, and transactions are all stored in a block. Blockchain is a decentralized storage that can build trust in the network and ensure that it is secure, transparent and trace able. In blockchain there is no central authority and is completely decentralized, which means that the transaction is verified by every member of the block chain network. This is a peer-to-peer (P2P) network where every transaction is cryptographically marked and is validated by checking it with each mining node in the network (English, Auer & Domingue, 2016). Each mining node in the network has a copy of the entire ledger (i.e., BC) which contains immutable records of all transactions that took place between the BC's various participants. No record can be changed once it has been saved on the BC without affecting all subsequent blocks in the BC (Cano-Benito, Cimmino & García-Castro 2019). It also provides detailed tracking and tracing of the block's trans-actions (Panarello et al. 2018). The main goal of the blockchain is protecting users' privacy and ensuring the authenticity of agreements by combining the benefits of P2P and cryptography technologies.

The blockchain is still in its infancy as a technology and the first stage is blockchain 1.0 (Tran, Babar & Boan 2021), (Hewa, Ylianttila & Liyanage 2021). This stage is represented by cryptocurrencies, with Bitcoin (Antonopoulos, 2017) being the most popular. Blockchain 2.0 enables the implementation of powerful Smart Contracts or executable programs and commands, gradually expanding its applications area and range. This phase could enlarge the blockchain application to include a variety of industries and allow them to collaborate among each other. The age of programmable society with blockchain of things will be the next generation of blockchain. Human philosophy and society form will be influenced by blockchain-related factors. Artificial intelligent apps such as Decentralized Application (Dapp), Decentralized Autonomous Organization (DAO) and Decentralized Autonomous Corporation (DAC) are starting to appear in the real world (Ante 2021). Blockchain technology combines and interoperates architectures, technologies, devices, and other relevant things to create high-quality goods and ser-vices for society and will become a strong tool for industry 4.0 soon (Ruta et al. 2017a) (Valiente, Rozas & Hassan 2014). A survey on blockchain technology applications and research challenges is summarized in this paper (Valiente et al. 2019) . Blockchain has a lot of attributes that make it useful. According to the UK government's office of research 'blockchain secures data records, lowers operational expenses, and increases

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

transaction transparency (Hector & Boris 2020). The following are some of the interesting aspects, benefits, and significance of blockchain:

Distributed Nature

Different users (nodes) on the network store the same blockchain data. Even if a node fails or loses data, other nodes in the network have a copy of the blockchain and can continue to update it. The blockchain can be recopied from other nodes by the impacted node. This property guarantees against data loss, record tampering, and cryptocurrency double spending.

Decentralized nature

Blockchain eliminates the needs for central authorities and intermediaries, it is more to trustless systems. Blockchain enables systems to be self-contained and independent of hazards that come with relying on middlemen and central authorities.

Data security and integrity

Blockchain is tamper-proof because each block relates to the hash of the previous block, so any data change in a block will be detected by the block hash. To be successful an attacker must change the block data for all computers on the network, which is essentially impossible in a big network.

Transparency and traceability

All activity and transactions can be examined and observed by everyone on the network since blockchain records are time-stamped and saved on all full nodes. All node`s activity and transaction can be tracked if its address is known. It is also a useful platform for auditing and public services because it is suitable for fraud detection.

Cost-cutting

Using blockchain saves a lot of money since it eliminates the need for intermediary systems. This is one of the main reasons why banks and businesses seek to integrate blockchain into their systems.

Verifiability

The validity of a record may be verified thanks to the cryptography used in the blockchain. This may be difficult to achieve in other database technology systems since it requires cryptographic technologies such as blockchain`s digital signature.

Anonymity

Asymmetric encryption methods are used to encrypt data on the blockchain. Data encryption on the blockchain provides data transaction security and lowers the risk of data loss or falsification. Every transaction data is

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

signed before it is transmitted to the network to identify the signatory`s identity. In the blockchain it is not required to reveal the real identity of the node linked to the participant.

Another important feature of blockchain is Smart Contract (SC). It is a block of programming code created in a specific language for different types of blockchain s that can be executed when a specific condition is satisfied. It minimizes transaction processing time and overhead cost while using blockchain networks.

Blockchains are classified into three types based on their uses and thresholds: public (permissionless) chain, private (permissioned) chain and hybrid chain.

A completely decentralized network is the public chain. Any node in a decentralized ledger can participate in the reading, writing, verification, and consensus procedures of the data on the chain, and get monetary rewards based on their contributions. Bitcoin is an example of a public chain.

A centralized blockchain is a private chain. The central authority controls the user access to the data on the chain, and the read permission can be selectively exposed to the public, mostly for internal data management or auditing of certain companies. The private chain is dedicated to small organizations or specialized enterprises.

Hybrid chain is a combination of public and private chains, it is partially distributed. Each block creation is determined by a group of pre-selected nodes. Other nodes in the system can only access the blockchain to handle the transaction, but they are not involved in the consensus process. Multiple organizations can join to create a consortium system for common goals by reaching a consensual agreement.

In blockchain there are consensus algorithms (proof-of-work, proof-of-stake, etc.) that are used to add a new block to the network, this algorithm checks and acts on the majority decision judgements.

In essence, blockchain technology`s mission is to replace the need for a centralized database and traditional setup with an autonomous access control mechanism. Consensus algorithms, distributed ledger, timestamps, Merkle Tree, and digital cryptography keys are among the computer skills and algorithms included in the blockchain. The blockchain combines several modern technologies, such as IoT, cloud computing, and data mining as a decentralized architecture. Apart from cryptocurrencies, blockchain has a wide range of uses, as well as numerous adoptions from a variety of nations and businesses. There is a potential that the world will become increasingly blockchain-based.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

3.2 CONSENSUS MECHANISMS

Consensus is the backbone of any blockchain because it facilitates decentralization of control when creating new records in the ledger. In other words, consensus is the process responsible for achieving agreement between the nodes in the ecosystem. The choice of consensus algorithm mainly depends on the blockchain in use and the ecosystem's requirements.

Nowadays, there is a plethora of consensus mechanisms (Nijsse and Litchfield, 2020), (Ferdous et al., 2020). In this deliverable we will cover the most widely used consensus mechanisms.

- **Proof-of-Work (PoW):** This mechanism relies on the proof that the nodes (i.e. miners) have spent adequate computational resources before proposing a value for acceptance by the network. This mechanism has proven to be successful against any collusion attacks and has been used by leading blockchains, such as: Bitcoin, Litecoin, Ethereum 1.0, Monero and many more.

- **Proof-of-Stake (PoS):** This mechanism relies on the idea that a node has an adequate stake in the system, meaning that the node has invested enough in the system and any malicious attempt by that node would not outweigh the benefits of staking in the network. PoS requires nodes to stake their tokens to become validators in the network. Validators are responsible for ordering transactions and creating new blocks so that all nodes can agree on the state of the network. PoS delivers many improvements to the PoW, such as: better energy efficiency, better scalability etc. PoS is used by Ethereum 2.0, NEO, Cardano and other.

- **Federated Byzantine consensus:** This mechanism allows the nodes to retain a group of publicly-trusted peers and propagate only those transactions that have been validated by the majority of trusted nodes. Hence, a Byzantine Agreement is reached when a certain minimum number of nodes (known as a quorum) agrees that the solution presented is correct, thereby validating a block and recording it on the blockchain. This mechanism is used by Ripple, Stellar etc.

- **Delegated Proof-of-Stake (DPoS):** This mechanism is very similar to PoS. The innovation it delivers is that each node in the system can delegate the validation of a transaction to the other nodes by voting. Essentially nodes are voting for "witnesses" and "delegates" with placing their tokens on the name of their candidate (those tokens are not spent, because they are only representing the position of stakeholder and remain his/her property). DPoS is used by BitShares, EOS, Lisk etc.

- **Proof-of-Authority (PoA):** This is a reputation-based mechanism that operates by knowing and utilizing the identities of the validators as

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

a stake on the network. Once the validators are known, identified and authorized, they are allowed to propose new blocks and record data on the blockchain. Such a mechanism enables companies and organizations to maintain privacy and security, whilst exploiting the blockchain functionalities. Famous implementations of PoA include: Kovan (Ethereum test network), VeChainThor, Kaleido etc.

- **Proof-of-Storage:** This mechanism allows outsourcing of storage capacity and facilitates efficient verification of the stored data integrity. This mechanism is based on the idea that a data fragment is stored by a node, which serves as a mean to participate in the system. Essentially, an encoded version of the data is shared for verification at any given point in time, while keeping the data fragments locally. Most famous implementation of the Proof-of-Storage mechanism includes FileCoin.

3.3 NON-FUNGIBLE TOKENS

Non-Fungible Tokens are cryptographic assets stored on a digital ledger (e.g., blockchain) with unique identification codes (e.g., hash value) and metadata. Unlike cryptocurrencies, NFTs are not exchangeable. NFTs can be used to represent items such as audio, videos, digital art, and any other digital files. NFTs of such digital files can then be tracked on a blockchain for implementing proof of ownership.

The most famous NFT example is that of CryptoKitties[3]. CryptoKitties are digital representations of cats on Ethereum. Each one of them has a unique identification and a value/price. Later on, NFTs got a huge public interest and applied to digital entertainment such as digital art[4], music[5], and sports[6]. Now blockchains like Ethereum, Flow, and Tezos have their own standards for NFTs.

ERC-721 is a known standardization to generate NFTs. It defines ownership, security, and metadata with a minimum interface. Since NFTs cannot store these digital assets on blockchain, the digital assets are stored on IPFS. The owner shares the corresponding URL in IPFS with the buyer on a smart contract. NFTs are based on unique cryptographic hash values created based on the asset. Because of the cryptographic (collision-resistant) hash functions, the linkage between the original asset and its NFT is only maintained if the original asset remains intact. If the original asset changes, even in a few bits, then its hash value changes and hence its linkage with the NFT is broken since the new hash no longer matches the one recorded in the NFT.

[3] https://www.cryptokitties.co
[4] https://www.theverge.com/2021/3/11/22325054/beeple-christies-nft-sale-cost- everydays-69-million
[5] https://www.rollingstone.com/pro/news/kings-of-leon-when-you-see-yourself- album-nft-crypto-1135192/
[6] https://nbatopshot.com

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

3.4 BLOCKCHAIN LAYER SOLUTIONS

Blockchains can also be layered to account for scalability, performance, energy-efficiency, sustainability, security and other trade-offs. In the following we provide a brief account of these approaches.

3.4.1 Layer-1 vs Layer-2

Layer-1 is the term that's used to describe the underlying main blockchain architecture. Layer-2, on the other hand, is an overlaying network that lies on top of the underlying blockchain. Consider Bitcoin and Lightning Network. Bitcoin is the layer-1 network, while the lightning network is layer-2.

3.4.1.1 Layer-2 Solutions:

State Channels

A state channel is a two-way communication channel between participants, which enables them to conduct interactions, which would typically occur on the blockchain, off the blockchain. Doing this helps in cutting down the waiting time since you are no longer dependent on a third party like a miner.

This is how a state channel works:

- ✓ A portion of the blockchain is sealed off via multi-signature or some sort of smart contract, which is pre-agreed by the participants.
- ✓ The participants can directly interact with each other without submitting anything to the miners.
- ✓ When the entire transaction set is over, the final state of the channel is added to the blockchain.

Bitcoin's Lightning Network and Ethereum's Raiden Network are the two most popular state channel solutions. Both utilize Hashed Timelock Contracts (HTLCs) to execute state channels. While Lightning Network allows participants to conduct many microtransactions in a limited time period, the Raiden will enable participants to run smart contracts through their channels as well.

Nested Blockchains

Currently, OmiseGO, an Ethereum-based dApp, is working on a nested blockchains solution called Plasma. The design principle of plasma is straightforward:

- The main, base blockchain is going to lay down the ground rules of this entire system. It will not directly take part in any operations unless it needs to resolve some disputes.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

- There will be multiple levels of blockchains sitting on top of the main chain. These levels will be connected to each other to form a parent-child chain connection. The parent chain delegates work amongst its child chains. The child chains then execute these actions and send the result back to the parent chain.
- Not only does this solution significantly reduce the load in the root chain, but, if executed properly, it will increase scalability exponentially.

Layer-2 Solution Pros

- The biggest pro is that it does not disarrange the underlying blockchain protocol.
- Layer-2 solutions like state channels, and particularly lightning network, to conduct multiple microtransactions without wasting time with miner verification and paying unnecessary transaction fees.

3.4.1.2 Layer-1 Solutions

What this essentially means is improving the base protocol itself to make the overall system more scalable. The two most common layer-1 solutions are:

- Consensus protocol changes.
- Sharding.

Consensus protocol changes

Many projects like Ethereum are moving on from older, clunkier consensus protocols like Proof-of-Work (PoW) to faster and less wasteful protocols like Proof-of-Stake (PoS). Bitcoin and Ethereum both use PoW, wherein miners solve cryptographically hard equations by using their computational power. While PoW is secure, the problem is that it can be very slow. Bitcoin only manages 7 transactions per second, while Ethereum can only manage 15-20 on a good day. Therefore, Ethereum is looking to change over from PoW to PoS (via the Casper protocol).

Sharding

Sharding is one of the most popular layer-1 scalability methods that multiple projects are currently working on. Instead of making a network sequentially work on each and every transaction, sharding will break these transaction sets into small data-sets called "shards." These shards can then be parallelly processed by the network

Layer-1 Solution Pros

The biggest pro is that there is no need to add anything on top of the existing architecture.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

A Major Problem with both Layer-1 and Layer-2

There are two significant issues with Layer-1 and Layer-2 scalability solutions. Firstly, there is a big problem with adding these solutions to already existing protocols. Ethereum and Bitcoin both have multi-billion-dollar market caps. Millions of dollars are traded every single day using these two cryptocurrencies. Therefore, it doesn't make sense to add unnecessary codes and complications to experiment with these protocols and play around with so much money.
Secondly, even if you create a protocol from scratch, which has these techniques built-in, they can still fail to solve the scalability trilemma.

The term "scalability trilemma" was coined by Ethereum founder Vitalik Buterin. It is a trade-off that blockchain projects must make when deciding on how to optimize their architecture, by balancing between three of the following properties — decentralization, security, and scalability. E.g., Bitcoin wants to optimize security and decentralization, which is why they end up compromising on scalability.

The solution is to build a protocol from scratch with these solutions built in. Plus, it should also be able to solve the scalability trilemma. Turing award winner Silvio Micali is building a project called "Algorand," which is trying to precisely do that. Algorand uses a consensus protocol called Pure Proof of Stake (PPoS).

During PPoS:

- The leader and selected verifiers (SV) are chosen from each step of the Byzantine Agreement.
- The computation cost a single user faces only involves generating and verifying signatures and simple counting operations.
- The cost is not dependent on the number of selected users for each block. This number is constant and unaffected by the size of the whole network.
- Increasing computational power directly improves performance, which makes Algorand perfectly scalable. This means that as the network increases in size, it sustains a high transaction rate without incurring extra costs.

Scalability is the biggest reason inhibiting the mainstream adoption of cryptocurrencies. To make sure that cryptocurrencies are scalable and fast enough for day-to-day transactions, we need protocols that have been built specifically to solve this problem. Therefore, projects like Algorand are critical, and we can only hope that other projects follow suit and provide a viable solution (Wallance, 2020).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

3.5 MAIN DLT NETWORKS

3.5.1 Bitcoin

Bitcoin is the first blockchain network that was developed in 2008 as an open-source and public DLT. Since then, the Bitcoin has set the pathway for thousands of blockchains that have been created using similar cryptographic techniques.

3.5.2 Ethereum

Ethereum is a decentralized, open-source blockchain that went live in 2015. It is one of the most famous representatives of the Blockchain 2.0. Essentially it delivered the smart contract functionality, and apart of the peer-to-peer payment functionality it integrated the business logic in the smart contracts. This traced the path for the development of thousands of decentralized applications, whose business logic is recorded and secured on the blockchain.

3.5.3 Tezos

Tezos is an open-source, PoS blockchain network that was released in 2018. Similarly, to Ethereum it allows the execution of peer-to-peer transactions and the deployment of smart contracts on its network. Due to its PoS consensus mechanism, this network is energy-efficient and has been chosen by many brands to also build their NFTs (e.g. Red Bull Racing, McLaren Racing).

3.5.4 EOS.IO

EOS.IO is another open-source blockchain network that was released in 2017, which was meant to provide means for development of decentralized applications. In comparison to other ledger technologies, EOS.IO introduced high-speed transactions while eliminating fees charged to users making the transactions.

3.5.5 IOTA

IOTA is an open-source distributed and centralized ledger that is designed for secure exchange of data in the Internet of Things. To achieve greater scalability, this blockchain network uses a directed acyclic graph to store transactions on its ledger. Because IOTA does not use the services of miners to record transactions on the network, transactions are issued without fees.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

4. SEMANTIC WEB AND BLOCKCHAIN

Blockchain technology is a decentralized peer to-peer architecture that includes privacy, security, permission, transparency, accountability, identity management, and trust. Although Blockchain technology offers immutability and trust, the huge number of current data and standards for decentralized web data distribution and processing should also be considered. One of the Semantic Web vision's major design principles is that data may be semantically annotated, disseminated anywhere online and by anyone, and that it should be simple to query and interlink the data without having to aggregate it all in one location. To get the most out of both technologies, Mikroyannidis, Third, and Domingue (2020) propose a Semantic Blockchain, which encourages interoperability between Blockchain networks and the Semantic Web. Semantic Blockchain also allows smart contracts on the blockchain to be mapped to contextual data about the relevant data. Semantic Blockchain is a realistic answer to the complicated issues of automation and constructive productivity, as mentioned in many research studies (Cano-Benito 2019, Hector 2020). In contrast to the present, heterogeneous Web, Blockchain provides the foundation for building a new Semantic Web. According to (Aswini, 2021), Semantic Blockchain is the application of Semantic Web principles to blockchain-based systems. The latter supports standard data formats and trade practices on the blockchain by utilizing the Resource Description Framework (RDF) and Linked Data standards. It also allows entities to be defined using URIs and resource definitions, as well as annotated with semantic information using Ontologies (e.g., RDFS, OWL). Furthermore, English, Auer and Domingue (2016) presented a special benefit of the Semantic Web for Blockchain practitioners: by utilizing ontologies to describe the meta-data of the blocks, they may run SPARQL queries to search across the blockchain's accessible data. Furthermore, Ruta et al. (2017a) and (2017b) discuss the application of semantics in a Blockchain to increase the scalability of the IoT domain. Moreover, Cai et al. (2021) introduced a solution incorporating Semantic Web technologies to create a smart contract manager for the Ethereum network that enables indexing and semantic-based discovery by using meta-data as semantic data in the Ethereum smart contracts to enable semantic-based resource discovery within Blockchain processes. It is obvious that the usage of blockchains may provide data integrity. Semantic knowledge and data mapping from other chains or smart contracts, however, can be done only with the Semantic Web methodologies and standards. These technologies can potentially work together to provide a trustworthy technical stack for the new Web evolution that ensures data integrity, promotes safe data exchange, and powers linked data insights. Semantic web technologies and Blockchain integration can advantage in the following:

- **Linking data:** Linking data is a well-known and well addressed topic in the Semantic Web culture. One of the advantages of RDF is its capacity to retrieve data from other datasets, resulting in a global comprehension of the information even though the many pieces are

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

stored throughout the network. Blockchain may be able to profit by linking block contents to external databases, or even other block contents on the same or other chain (Mohsen, Arel and Elbahnasy, 2020).

- **Interoperability:** For both meta-data and information, the Blockchain becomes interoperable by depending on semantic web technologies and conventional ontologies. Interoperability guarantees that an information system can connect with other networks, such as other blockchains, databases, or services, in an easy way (Belchior et al. 2022)
- **Blockchain data and metadata validation:** Validation and knowledge reasoning are achieved by annotating data using Ontologies and organizing it with RDF (Valiente, Rozas and Hassan, 2021).
- **Searching over Blockchain:** Researchers can query the blockchain using SPARQL if the Blockchain employs ontologies and RDF to describe its meta-data and metadata.

4.1 PRELIMINARIES

Semantic Web tools and languages pursue the goal of achieving full interoperability of computer systems, promoting common data formats, exchanging protocols on the web, sharing, and reusing data across applications and across enterprise and community boundaries. In the Semantic Web vision of the internet, software agents are enabled to query and manipulate information on behalf of human agents by means of machine-readable data carrying explicit meaning, hence they can automatically process information thanks to appropriate representation languages that express meaning with formally defined semantics. To reach this objective, the Word Wide Web Consortium (W3C) conceived the Web Ontology Language (OWL) Hitzler (2009), a family of knowledge representation languages relying on Description Logics (DLs) (Baader et al. 2017), as the standard language for representing Semantic Web ontologies.

An ontology is a formal description of a domain of interest carried out by combining three basic syntactic categories: entities, expressions, and axioms. These form the logical part of ontologies, namely what ontologies can express and the type of inferences that can be drawn (Oberle et al. 2009, Hofweber et al. 2018). Ontologies can also be combined to describe more complex domains. A specific type of ontologies, namely upper-level or foundational ontologies (Oberle et al. 2009), is designed to model higher level and independent categories concerning the real world. Foundational ontologies provide general terms that are used to connect domain-specific ontologies (also called lower-level ontologies), allowing one to reach a broader semantic interoperability. For instance, the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) (Gangemi et al. 2002) is designed to capture the ontological categories underlying natural language and human common sense, whereas the CIDOC Conceptual Reference Model (CIDOC-CRM) (Doerr

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

et al. 2007) is conceived to model concepts and information in the ambit of cultural heritage and museum documentation.

OWL, currently in version 2.1, provides users with constructs useful for designing ontologies for real world domains that are available neither in the basic Semantic Web model Resource Description Framework (RDF) (Manola and Miller 2004), nor in the basic Semantic Web language RDF Schema (RDFS) (World Wide Web Consortium. Linked data platform 1.0, 2015.) the latter being an extension of RDF admitting taxonomies and primitives to denote range and domain of relations and subsumption axioms. As RDF, OWL exploits the notion of RDF triples (Cyganiak et al. 2014), which are ways of connecting entities or resources. The term resource is used to indicate any physical or digital entity described via Semantic Web technologies. Resources are identified by an Internationalized Resource Identifier (IRI) (Duerst and Suignard 2005), an extension of the Uniform Resource Identifiers (URIs) (DuCharme 2011) with internalization features. Resources should be also deferenceable (World Wide Web Consortium. Linked data platform 1.0, 2015.) via the HTTP protocol (Fielding and Reschke 2014) using content negotiation and, in the case of digital resources, they must be accessible via their IRI.

Facts and relations involving resources are stated by means of triples consisting of the subject resource a property, labelled by an IRI, and an object which may be either a resource or a concrete datum (data type value such as strings and numbers).

Resources are individuals (actors), properties (actions), and classes (sets of actors with common features). Properties are of two types: object-properties and datatype-properties. Object-properties relate pairs of individuals, whereas datatype-properties relate individuals with some data type values. Triples can be combined to build knowledge. Specifically, RDF triples are embedded in sets called RDF graphs (Klyne and Caroll 2004), which may be either named graphs, i.e., RDF graphs with associated names defined as IRIs, or unnamed graphs; in their turn RDF graphs can be collected in RDF datasets (Cyganiak et al. 2014), containing exactly one unnamed graph, called default graph, and zero or more named graphs. More details on these notions can be found in (W3C SPARQL 1.1 Query Language, 2013.).

The expressivity of the OWL language is extendable by the introduction of rules of the Semantic Web Rule Language (SWRL) (W3C SWRL: A Semantic Web Rule Language Combining OWL and RuleML, 2004.): SWRL rules are Horn-like rules consisting of an implication between an antecedent (body) and a consequent (head), intending that whenever the conditions specified in the antecedent hold, the conditions specified in the consequent must hold too. The reader is referred to Allemang and Hendler (2011) for details on the SWRL language.

Semantic Web knowledge is modifiable and retrievable by means of the SPARQL Protocol and RDF Query Language (SPARQL) (W3C. SPARQL 1.1 Query

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Language, 2013.), promoted by the W3C as the standard query protocol for RDF datasets. Mostly like SQL, the SPARQL language is a declarative language conceived to query and perform modifications on RDF graphs by executing SPARQL queries. SPARQL queries are constituted by a head and a body: the head comprises a modifier identifying the corresponding type of query (i.e., ASK, SELECT, and CONSTRUCT), whereas the body consists of an RDF graph pattern. Most notably, SPARQL CONSTRUCT queries allow one to retrieve information from a queried dataset and express it as new RDF triples. A detailed overview of SPARQL can be found in DuCharme (2011).

4.2 ONTOLOGIES FOR THE BLOCKCHAIN MARKET

OASIS (Ontology for Agents, Systems, and Integration of Services) (Cantone et al. 2019, 2021) is a very recent foundational OWL 2 ontology modelling multi-agent systems by representing agent's trough their behaviors, namely, purposes, goals, responsibilities, services provided, information about the world they observe and maintain, and their interactions with other collaborating agents. Additionally, OASIS models information concerning executions and assignment of tasks, restrictions on them, and constraints used to establish responsibilities and authorizations among agents. In Cantone et al. (2019) the ontology has been exploited to define an ontology-based protocol for the Internet of Agents (IoA) and to realize a transparent communication and information exchange system among agents via Semantic Web technologies. The resulting protocol is founded on the exchange of OASIS fragments, each consisting of a RDF description of a request that is checked, by means of suitably constructed queries, against the corresponding RDF description of the agent behavior selected to satisfy it. OASIS models agents by representing their behaviors which are publicly exposed. By exposing behaviors, agents report to the communication pairs the set of operations that they can perform and, eventually, the type of data required to execute them and the expected output. Representing agent behaviors has many advantages since it permits abstracting from implementation details thus making the task of discovering agents extremely transparent and automatic. Agents may join a collaborative environment in a plug-and-play way since there is no need for third party interventions. Therefore, users can freely choose products and services according to their needs since provision methods and supply chains are clearly described and represented.

Moreover, by means of Semantic Web technologies and of automated reasoners, data provides machine-understandable information which can be processed, integrated, and exchanged by any type of agent at a higher level. Data consistency can be easily verified, and information can be inferred and retrieved by exploiting what is already provided by the knowledge base. Representation of agents and their interactions in OASIS is carried out along three main steps.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

The first step consists in defining the agent behavior template: templates are high-level descriptions of behaviors of abstract agents that can be implemented to denote more specific and concrete behaviors of real agents. For example, a template may be designed for agents whose behavior consists in selling and shipping products to buyers, and it may be implemented by an apple seller that ships its products using the Fedex courier. Templates are useful to guide developers to define the most suitable representation of their agents.

The second step consists in representing the agent behavior either by implementing a template or by denying it from scratch. Agent behaviors are represented by the goals to achieve, in their turn goals are related with their constitutional elements, namely tasks. Tasks represent atomic operations that agents execute and are described by actions to perform. Actions are drawn from shared and common vocabulary and can be simple or composed, eventually associated with requested input parameters and expected outputs.

In the third step, in OASIS actions performed by agents are associated with the behavior are associated with the behaviors that generate them. To describe such association, OASIS introduces plans, which represent the will of agents to activate some agent behaviors or to get some activities performed. Plans in their turn are associated with their executions that also provide information about the obtained output.

Semantic description of goods and commercial offerings has been recognized as a crucial task for eCommerce (Goldmann 2021, Stonebraker and Hellerstein 2021, Jovanovic and Bagheri 2016) as it enables the implementation of semantic search engines (Guha et al. 2003) to find out items in a very specific range. GoodRelations (Hepp 2008) is an OWL vocabulary describing offerings on products or services, legal entities involved in them, prices, offering terms and conditions. Well-known content management tools as Joomla, osCommerce, and Drupal support publishing data with the GoodRelations ontology (Ashraf et al. 2014). GoodRelations has been integrated in schema.org (Patel-Schneider 2014), a general-purpose vocabulary largely used for tagging web page contents by means of Resource Description Framework in Attributes (RDFa) (World Wide Web Consortium. RDFa Primer: Bridging the human and data webs, October 2008.). Thus, offerings described using GoodRelations in conjunction with schema.org and published via RDFa are recognized by major search engines such as Google, Yahoo, and Bing, which use RDFa to enhance the appearance of individual search results with a structured description (Birbeck 2010).

The core class of the GoodRelation vocabulary allows one to represent an Offering. An offering is an announcement of an agent providing a certain business function, which is one of sell, lease out, maintain, repair, provide service, dispose, and buy, for a certain product or service instance to a particular target audience and under commercial conditions.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

A business entity, i.e., a legal agent, can create such an offering or seek for someone else providing goods and terms under conditions.

The BLONDiE ontology (Ugarte Rojas 2017) aims to achieve such a goal by modelling information concerning the two currently most relevant cryptocurrencies, namely Bitcoin and Ethereum, potentially covering and easily extendable to every blockchain available. The ontology tries to answer (not only to) the following main competency questions:

1. Who mined a specific block?
2. What is the height of a specific block?
3. How many transactions are written in a block?
4. Is a transaction confirmed?
5. How many total coins were transferred on a block?

BLONDiE, currently at version 0.4, is an OWL ontology comprehending 21 classes, 11 object properties and 50 data-properties.

4.3 GRAPH DATABASE & TRIPLESTORES

Graph databases are usually developed with the scope of connecting data between each other. More specifically, graph databases were designed to tread the relationships between data. Graph databases use topographical data models to store the data. The main reasons that led to the usage of a graph database are summarized below:

- **Easy creation and maintenance:** Graph databases are easily deployed and maintained.
- **CRUD (Create, Read, Update, Delete) enabled:** As in relational databases, graph databases support CRUD operations.
- **Scalable and Flexible:** Graph databases can scale quickly if required and in most of the cases faster than traditional databases, developing relationships between new and existing nodes.
- **Easy Querying:** Querying over graph databases are based on direct or transitive relationships instead on complex queries with foreign keys, nested queries etc.
- **All OS Compatible:** Graph databases do not require specific OS to operate, which make them more flexible choice for each case.
- There are few different approaches of what constitutes a graph database. One approach is **the property graph model** where nodes and relationships store and organize the data and properties describe them. In more details, the main components of the property graph model are described below:
- **Nodes:** Used as entities in the graph. They are able to hold any number of attributes (key-value pairs) the so-called properties.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Nodes can be tagged with labels, that server to attach metadata to nodes such as index or constraint information.

- **Relationships:** Provide the names, direction, and semantically relevant information for the connection between two node entities. Any relationship has a direction, a type, a start node (entity) and destination node. In addition, relationships could have properties (e.g., weight costs, distance etc.)

Another type of graph databases is those which store the data as triples referred also as triplestores. Triples use the (subject, predicate, object) format to describe the relationship between two entities. In the following paragraphs the top ten graph databases based on G2 scores are presented

Neo4j:

- Available in both open source and commercial licenses for enterprise levels of security.
- Leading native graph database and graph platform
- Cypher is Neo4j's query language.
- Can operate on Apache Spark and Gremlin.
- Offers a complete graph platform (Neo4jBloom, No4jETL, Kettle, Neo4j Browser)

ArangoDB:

- Fast growing native multi-model NoSQL database
- Combines the power of graphs, with JSON documents and key-value store
- It combines one database, one query language and three data models.

OrientDB:

- First multi-model distributed DBMS (Database Management System) with True Graph Engine
- Manages relationships without using JOINs, but direct pointers
- Allows constant performance on traversing relationships

Dgraph:

- World's most advanced GraphQL database
- Built for performance and scalability.
- Jepsen tested (best performance, returns millisecond query responses over terabytes on data)
- Ideal for a range of use cases (e.g., fraud detection, customer 360)

Amazon Neptune:

- Fast, reliable, and fully managed graph database.
- Supports popular graph models Property Graph and W3C's RDF

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

- Supports various query languages including TinkerPop Gremlin, SPARQL.
- Neptune is secure with support for encryption at rest.

FlockDB:

- Simple graph database
- Scales horizontally
- Designed for on-line, low-latency, high throughput environments

DataStax:

- Experiences partner in on-premises, hybrid, and multi-cloud deployments
- Offers a suite of distributed data management products and cloud services

Cassandra:

- Its data model offers the convenience of column indexes
- Strong support for denormalization and materialized vies
- Powerful built-in caching

Titan:

- Scalable graph database optimized for storing and querying graphs
- Support thousands of concurrent users executing complex graph traversals.

GraphDB:

- Allows linking of diverse data, index it for semantic search and enrich via text analysis to build big knowledge graphs.

Open-source plugins for connectivity with FTS engines, NoSQL and EQL databases, virtualization, community and commercial support.

4.3.1 AI-Based Knowledge Extraction

For many years, knowledge extraction and reasoning over semantically annotated graphs has been built upon two different approaches, each one holding its own advantages and disadvantages.

On one hand, rule-based reasoners (aka symbolic reasoners) can be considered as systems capable of inferring new symbolic and logical rules. However, rule-based reasoners suffer from two major drawbacks. First, they are affected by noisy data and the inferred triples are dependent on the veracity of the input triples, affected by several noise types. Secondly, they lack support for approximate reasoning, targeting complete inference that is time consuming. On the contrary, in interactive applications, it is preferable to provide the user with an incomplete set of inferred triples in a timely manner rather than a delayed complete inference.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

These two drawbacks triggered the second approach that is based on the research of deep learning techniques (aka neural reasoners). Deep learning has achieved superior performance on many tasks and has reached unprecedented impact across research communities. Moreover, studies have indicated that deep learning models can model the implicit correlations inside data. This can be shown in various fields such as image classification in computer vision (He et al. 2020, He et al. 2016), language modelling in natural language processing (Devlin et al. 2018), and link prediction in networks (You at al. 2020). Unfortunately, deep learning-based reasoners suffer from the lack of explain ability, which is a problem that is common for all deep learning approaches in general. As stated by Makni et al. (2020), the lack of explain ability translates into the inability to provide the derivation of the inferred triples.

4.4 STATE OF THE ART CHALLENGES

4.4.1 Technology challenges

The ONTOCHAIN project's main goal is to provide a software ecosystem that integrates Blockchain and Semantic Web technologies in a way that addresses the current problems of the Internet. Hence, first and foremost it is necessary to address a number of technological challenges.

4.4.2 Decentralization of heterogeneous components

Transferring each diverse component to operate effectively and safely in a decentralized manner, connected to a Blockchain, will involve modifications in addition to the work of changing their interfaces and semantics.

4.4.3 Open design and flexible

As Blockchain technology changes rapidly, the innovator must make the balance between granularity and how much data is kept on-chain vs. performance, that may change as future Blockchain protocols emerge.

4.4.4 Formal logic proofs

Another technical problem that has arisen is how to extract a new fact from a collection oh know truths in a transparent manner. Formal logic may be expressed using a variety of languages, such as the OASIS web ontology language. Lite, Description Logic (DL), and Full are the three degrees of representation in OWL. While formal rules are easy to calculate in Lite and taker longer in DL, the full version cannot utilize formal proofs since the reasoning program can cycle forever, as Kurt Gödel's theorems demonstrate.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

5. STATE OF THE ART OF DIFFERENT TECHNOLOGIES EMPLOYED IN ONTOCHAIN

5.1 DECENTRALIZED REPUTATION SYSTEMS

5.1.1 Reputation Systems

Reputation represents collective information about the trustworthiness of users and retailers in the online marketplace. The reputation of a given retailer or user at the marketplace is computed as the sum or average of the feedback ratings assigned to retailers by their buyers based on the past transactions (Josang, Ismail and Boyd, 2007), (Hendrikx, Bubendorfer and Chard, 2015). In the marketplace the feedback ratings can be represented on the scale of {0, -1, 1} (eBay marketplace), 0-5 star (Amazon marketplace) that is assigned to each attribute describing the performance of a retailer. The attributes can be the information about whether the retailer has delivered the product on time, whether the product is the same as what is listed on the retailer's page and quality of services etc. Figure 2 shows the flow of events that take place in the marketplace when a consumer submits the purchase order to a particular retailer. When the product is delivered to the consumer, the online marketplace asks the consumer for the feedback against her recent interaction with the retailer. The user reports a feedback rating for the retailer and the marketplace then adds this trust score to the aggregated reputation of the retailer and displays this value on the webpage designated for the retailer.

Reputation systems can assist consumers to evaluate the trustworthiness of other retailers or users (consumer to consumer marketplaces or P2P) before making the transaction with a retailer or a user. This would also boost the sales for a particular retailer as well as people's trust on the marketplace if the reputation is well conceived and accurate in its functionality. The reputation system can be implemented as a centralized system or a distributed system depending on the requirement. In electronic marketplaces, the reputation systems are the centralized ones, where all the data regarding the rating history of the user is held by the central system. This is the normal practice that is being used in popular reputation systems (Amazon, eBay, Airbnb, uber etc.). On the other hand, in P2P network, the reputation systems can be distributed (Kamvar et al. 2003, Gupta et al. 2003, Walsh and Sirer 2006) where user ratings are retained within the peers and used on-demand upon request from other peers.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

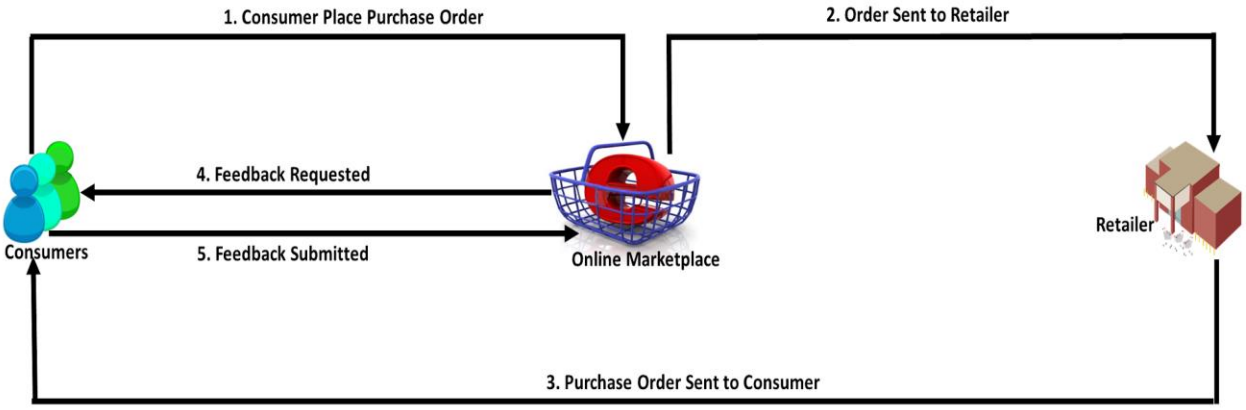

FIGURE 2: TRANSACTION AND REPUTATION FLOW WORK FOR THE ONLINE

5.1.2 Decentralized Reputation Management

Pavlov, Rosenschein, and Topol (2004) proposed a set of protocols for decentralized reputation aggregation. The protocols operate under the two adversarial model: honest but curious where participants are honest in their feedback submission but try to learn information from the available public information, and malicious model where participants not only learn private information of participants but would also disrupt functionality of aggregation system by providing out-of-range feedback values. The motivation of malicious users is to maliciously increase or decrease the reputation of some victim providers. The aggregation mechanism is based on secure multi-party computation, private secret sharing, and discrete log commitment. Reputation in this setup is computed as the inner product of a vector of reputation ratings given by voters and a weight of ratings of trust a requester places in each voter.

Kinateder and Pearson (2003) present a privacy-preserving framework by implementing a trustworthy mechanism for shaping and gathering sensitive recommendations within the node's Trusted Platform (TP). The trusted Platform at the participating nodes has the ability to prove its legitimacy without disclosing its real identity. In this setup a trusted agent creates suggestions and decides what is necessary to send out to other nodes anonymously. This, or another trusted agent, may be used to create queries requesting advice from other nodes.

Bag et al. (2018) presented PrivRep, a privacy-aware decentralised and customized reputation system for electronic marketplaces. The systems compute the trust score for business entities in a completely decentralized and anonymous way. The aggregate score of business entities is computed without revealing the identities or trusted scores of the participants. PrivRep's architecture contains raters, a marketplace, and a public bulletin board (PBB). The PBB can be implemented as the

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

blockchain setup as proposed in the recent papers from the authors of PrivRep and PrivBox (Azad et al., 2018).

Dou et al. (2019) propose a privacy-preserving distributed trust management system for the Intercloud setup. The silent feature of the system is that it can still provide reputation aggregation or reputation evaluation even if some the participants of the system go online during the aggregation or feedback process.

Lu et al. (2018) proposed a blockchain based privacy-preserving reputation model for the vehicular ad hoc networks. The participating vehicles can anonymously submit the behavior of driver and conditions of roads. The group of users then collaborate to evaluate the trustworthiness of vehicle nodes engaged in providing the feedback. The driver anonymity can be anonymized using the background information or correlating the information from different sources. To have a complete anonymous system, Azad et al. have proposed a decentralized reputation system where multiple vehicles could report the feedback in an encrypted form which could only be revealed as an aggregate. The participants malicious or honest but curious would not be able to learn anything about vehicle feedback or their position on the road. The system also has the inherent property of providing the aggregation process even if some of the vehicle nodes go online during the process. For the Crowdsourcing Based Reputation Systems, several approaches have been proposed to guarantee the privacy of users while computing aggregate statistics over their shared values. Yang et.al (2015) identify many security and privacy challenges that are essential for the design of a privacy preserving crowdsourcing system. Rashidi et al. (2017) proposed a DroidNet, a framework that assists mobile users to have feedback from other users about privacy-related permissions of applications. The objective is to identify malicious apps. However, the DroidNet framework itself can easily learn about the user's apps usage. Jin and Zhang (2018) proposed a novel framework to select spectrum-sensing participants in a privacy-preserving way. The framework is based on the semantics of differential privacy (Dwork, 2006) and ensures the privacy of location privacy and truthfulness. Zhang et al. (2018) also adopted differential privacy under the non-trusted server setup to ensure the privacy of participants in the crowdsourcing system. However, adding noise to data where the accurate result is necessary is not a desirable choice.

Erlingsson, Pihur and Korolova (2014) presented RAPPOR (Privacy-Preserving Aggregatable Randomized Response) for collecting statistics from clients while providing strong semantics of privacy-preservation using randomize response generation. RAPPOR collects a user's feedback or values about the set of strings using Bloom filters (Broder, 2004) with strong differential privacy guarantees. Polat et al. (2003) proposed a collaborative filtering solution that randomized the user's responses using Randomized Perturbation techniques with the inclusion of the noise. Erkin et al. (2012) proposed the system for generating the recommendation by encrypting the user's responses (rating for certain products or

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

objects) in the homomorphism-based cryptographic system. Azad et al. (2017) proposed a collaborative system that considers the encrypted feedback and weights of providers for computing the reputation of users in the respective content provider. However, the system is not completely decentralized as it depends on the trusted setup for the protecting assigned weights of raters. Wang et al. (2018) proposed a distributed agent-based privacy-preserving framework, called DADP, which consists of multiple agents that handle the user responses before relaying them to the untrusted server. Melis, Danezis, and Cristofaro (2015) proposed a decentralized system for enabling users to participate in providing the feedback for different applications. (Gibbs & Boneh, 2017) proposed a Prio system that consists of clients who hold the private data value and a small set of servers for computing the statistical function over the values reported by the clients.

The privacy of the client is purely dependent on the honesty of the servers. (Primault et al., 2019) proposed a Private Data Donor (PDD) platform for aggregating the web query results in a decentralized and privacy-preserving way. (Bonawitz et al., 2017) proposed a scheme for aggregating the values represented as a vector. The scheme ensures the privacy and security of participants under the honest-but-curious and malicious adversaries. (Halevi, Lindell & Pinkas, 2011) proposed an aggregation scheme based on the homomorphic cryptosystem that evaluates the mathematical function securely and privately. However, the scheme requires PKI. (Miao et al., 2019) proposed a framework that performs a weighted aggregation over the user's encrypted data. The framework employs a homomorphic cryptosystem that has high accuracy in aggregation as well as protects the privacy of users. However, weights in this scheme are sent directly to participants. Melis, Danezis, and Cristofaro (2016) proposed efficient cryptographic methods for the private aggregation of the large data stream. The data aggregation is performed in a privacy-preserving way using data sketches, instead of the raw data inputs. (Liu et al., 2019) proposed an anonymous reputation system for the retail market that ensures privacy of consumers by using blockchain technology. The system protects the real identity of the user and his review using the anonymization approach; however, the private information of users can be deanonymized using some background information e.g. the buying history of the users. (Yang et al., 2019) proposed a blockchain-based decentralized anonymous credential system that exchanges the list of users blacklisted by the particular user in a privacy-preserving way. The system utilizes the tally like system for the sharing of the blacklist. (Wang & Singh, 2010) proposed a trust and reputation model for the multiagent systems that use how agents in the system would produce the trust score from the evidence of their direct interactions. The system does not provide any discussion on how the privacy of participating agents is protected. In our work, we estimated the trustworthiness of the nodes (objects, content creators) while also protecting the privacy of the participant's feedback. A privacy-preserving solution is proposed for the spatial crowdsourcing (Yuan et

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

al., 2019). The scheme ensures privacy in two aspects: firstly, protection on the location of users in the crowdsource group, and secondly, the content of tasks is protected against the server and other users in the crowdsource group. To protect the location privacy the authors, divide the location into grids and encrypt the grids as the code. For this purpose, the authors use attribute-based encryption and symmetric-key encryption. (Wu, Wang & Xue, 2019) proposed a data aggregation scheme using the bilinear pairing and homomorphic encryption. However, the scheme requires a third-party system i.e., a Fog computing server to ensure the privacy of workers in the network. (Fredrikson et al., 2014) proposed to protect the privacy of patients and analyze the risk to the health of the patient using differential privacy with different privacy budgets. (Kim, Kim & Jang, 2018) presented an effort to address the challenge of protecting the privacy of health data streams emerging from smart devices. However, the aggregator or collector is a central component usually hosted by the healthcare service provider. (Zheng et al., 2017) proposed crowdsensing methods that utilize the design choice of trust discovery. The design has inherent properties of privacy-protection of participants and have reasonably improved bandwidth and computation requirements for the participating users. However, the proposed systems require the trusted server for the handling of data and computation of results.

5.2 DATA PROVENANCE

One of the major advantages of blockchain has been its ability to store data in a tamper-proof manner as part of a consensus blockchain. Consequently, this ability leads to trustworthiness of data stored in the blockchain ledger. Therefore, blockchain has been used in a number of diverse scenarios as a trustworthy data store which can guarantee integrity of the data. Similarly, the use of blockchain is prevalent in applications which require a trustworthy storage for provenance data. A large proportion of such applications are supply chain systems where trustworthy provenance data is crucial to traceability across a distributed trustless environment. Through our search queries, a significant number of papers were related to use of blockchain to maintain data provenance within supply chain systems. In this respect, (Pham, Adamopoulos, and Tait, 2019) and (Shwetha and Prabodh, 2021) are recent examples of such research.

We consider the use of blockchain to store provenance of reputation data of external (off-chain) services in a tamper-proof manner. In this respect, our focus is on existing literature which is particularly focused at the challenge of managing provenance of data originating from off-chain to achieve its trustworthiness. Therefore, we present a critical review of relevant existing efforts below.

With the increased use of blockchain technology, a number of efforts have been made to utilise the immutable property of blockchain ledger to aid recording provenance in a tamper-proof manner. ProvChain (Liang,

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

2017) represents one such effort where authors use blockchain to store cloud data provenance i.e. metadata about cloud data objects. The authors propose concept of provenance data receipts which are based on blockchain constructs such as transaction ID and block number, and are used a proof of validity of the provenance-related information. Authors utilise hashed user IDs to achieve anonymous storage of provenance of cloud data objects.

As illustrated by Videnov et al. (2019), provenance of off-chain information covers two aspects i.e. provenance of data when it is originated and provenance from the stage where it enters blockchain ecosystem. The authors presented a blockchain based system to explore challenges in maintaining provenance especially across different domains. Among others, the authors highlight challenges such as the oracle problem, the on-chain vs off-chain storage of provenance data, and heterogeneity of provenance data across different components of a system. These challenges are noteworthy and they will be taken into account towards an effective implementation of a decentralised reputation mechanism for the ONTOCHAIN ecosystem.

Lemieux et al. (2017) presented a novel framework for evaluating the capability of innovative blockchain-based systems to deliver trustworthy recordkeeping based on archival science (an ancient science aimed at the long-term preservation of authentic records). Author presented a blockchain-based reference architecture to preserve completeness, consistency, and naturalness of archival records. Naturalness refers to events that are expected to occur as part of daily routine and not caused purposefully. The reference architecture presented is generic and does not address new details with respect to data modelling and management.

The authors (Nasikas, 2018) present an accountable method for data storage and processing. Authors are especially focused on big data applications. The authors used a public blockchain-based auditing system which keeps a tamper proof log of actions performed by participants. However, authors assume dataset owners to be responsible for the quality, availability and security of their datasets. Furthermore, authors used off-chain datasets to enable data owners /controllers to be able to manage security of their datasets. Authors specifically consider use-cases where data controllers are collecting data with privacy concerns and therefore require additional measures for their security.

Ramachandran et al. (2017) represents another effort to use blockchain's ability to provide tamper-proof storage to record data provenance. Specifically, the authors focus on the challenge of verifying credibility of scientific experimentation results by recording and maintaining provenance of such data. DataProv system developed by the authors uses Ethereum smart contracts alongside a voting system to record changes to data on the blockchain. The authors also implemented automated verification scripts for the scientific data which enables maintaining a consistent version of the data.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

As illustrated by the authors (Caldorelli, 2020), too often, the words bitcoin and blockchain are confused, and it is evident that most of the papers address characteristics that strictly belong to Bitcoin, rather than to regular blockchains. Furthermore, the literature neglects that when implemented in the real world, smart contracts need oracles to operate. This paper investigates the roles of oracles in real-world applications. Oracles are the only means of communication for blockchain with the real world, and unlike blockchain nodes, they are centralized and exposed to tampering and manipulations. The risk of oracles being compromised and feeding the blockchain with false information is called the "oracle problem". The oracle problem biases all real-world applications, but its impact varies according to the application itself. The most promising and discussed smart contract applications, such as IPR protection, energy production, healthcare, supply chain management, academic transcript, and legal contracts, are thus analyzed. The analysis provided in this study supports the view that the oracle problem inevitably affects real-world applications. However, the impact is different, and it strictly depends on the trustworthiness of the system in which it is implemented.

The authors (Lopez-Pimentel et al., 2020) describe an audit mechanism that contains three significant parts: a) a supply chain architecture; b) a server blockchain interface; and c) a blockchain. So, the general proposal consists in saving all the events of part a) within a blockchain in a hashed way. The major contribution is a solution where the blockchain and the Server Interface can be implemented as an extension in systems requiring an audit characteristic in data provenance and traceability. The authors do not address the `oracle problem'. The provenance of off-chain data is acknowledged but no specific measures to address this provenance is not presented.

Blockchain technology has also been used to assure the privacy of users in an IoT network (Moin et al., 2019). Chen et al. (2019) designed a blockchain-based model to protect the privacy of participants in the big data environment. The system is more generally designed for protecting raw data but in our case, we protect the user's data while still performing some meaningful analytics over the encrypted data without actually decrypting it. Gan et al. (2019) proposed a privacy-preservation model for task allocation in a crowd-sourced environment. The privacy of IoT nodes which allocate jobs to others is protected by the means of task division and hiding the social network of IoT nodes. Fortino et al. (2019) designed a blockchain-based model to distribute the reputation score among nodes in a distributed IoT network. The proposed approach computes the reputation of each node in the network and then develops the collaborative network among nodes for the network-wide view about the trustworthiness of nodes in the network as a whole. Tang et al. (2019) proposed a protocol named IoT Passport that enables IoT devices from a different platform to collaborate with each other using the blockchain system. In this setup, the interaction between devices is

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

signed with a digital signature and recorded in the tamper-proof blockchain. A three-player game model is proposed by Tian et al. (2019) that protects private information and friendship network of devices and users in the context of the connected social Internet of Things.

Also, regarding ontologies for representing data provenance, the PROV Ontology (PROV-O) W3C recommendation[7] expresses the PROV Data Model using the OWL2 Web Ontology Language (OWL2). It provides a set of classes, properties, and restrictions that can be used to represent and interchange provenance information generated in different systems and under different contexts. It can also be specialized to create new classes and properties to model provenance information for different applications and domains.

Moreover, providing Records of Personal Data Processing Activities (ROPAs) is currently a mandate of the General Data Protection Regulation (GDPR) Regulation (EU) 2016/679) (EU 201657), which rules ROPAs in Article 30. ROPAs should not be independent and isolated pieces of information. They should be reliable sources of information, linked, available for intelligent knowledge extraction. While there are many semantic models that focus on GDPR concepts, there are currently no ontologies that model GDPR expert professional knowledge with a focus on the ROPA maintenance and management required by data controllers and supervisors of the records. For extensive accounts of data protection related ontologies see Pandit (2020) and Esteves & Rodríguez-Doncel (2021).

Legend:

- ✗ No expert involvement | Not publicly available
- ☒ Limited expert involvement
- ☑ Extensive expert involvement | Publicly available

	Purpose	Expert Knowledge	Formality| Expressivity	Publicly available
SPECIAL usage policy language	**"This document specifies the SPECIAL usage policy language, which can be used to express both the data subjects' consent and the data usage policies of data controllers in formal terms, understandable by a computer, so as to automatically verify that the usage of personal data complies with data subjects' consent".**	✗	OWL2	✓

[7] https://www.w3.org/TR/prov-o/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Data Privacy Vocabulary	"The DPV is a vocabulary (terms) and an ontology (relationships) serialised using semantic-web standards to represent concepts associated with privacy and data protection, primarily derived from GDPR. It enables representation of which personal data categories are undergoing a what kind of processing by a specific data controller and/or transferred to some recipient for a particular purpose, based on a specific legal basis (e.g., consent, or other legal grounds such as legitimate interest, etc.), with specified technical and organisational measures and restrictions (e.g., storage locations and storage durations) in place".	✗	RDF/OWL	✓
Policy Log Vocabulary	"This document specifies splog, a vocabulary to log data processing and sharing events that should comply with a given consent provided by a data subject. We also model the consent actions related to consent giving and revocation."	✗	RDF/OWL	✓
DVP-GDPR	"he Data Privacy Vocabulary (DPV) provides terms (classes and properties) to describe and represent information related to processing of personal data. This extension extends the DPV and provides concepts specific to the obligations and requirements of the General Data Protection Regulation (GDPR). More specifically, it provides a taxonomy of legal bases and rights as defined within the GDPR."	✗	RDF/OWL	✓
GDPRov	"GDPRov (pronounced GDPR-Prov) is a linked data ontology for expressing provenance of consent and data lifecycles with a view towards documenting compliance".	✗	OWL2	✓
GConsent	"The ontology is based on an analysis of modelling metadata requirements related to the consent lifecycle for GDPR compliance. It allows modelling and representation of information related to compliance in an extensible and comprehensive manner."	✗	OWL2	✓
GDPRtEXT	"The GDPRtEXT ontology aims to provide a way to refer and use concepts defined by the General Data Protection Regulation	✗	RDF/OWL(SKOS)	✓

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

	(GDPR).”			
Data Protection Ontology	“...the ontology will constitute the knowledge base from which the concepts to annotate the workflow model are extracted. Such an approach can provide benefits for a number of stakeholders: – data controllers would have a clearer view of their duties with respect to data protection in the context of their business; – the auditors would have a first-look model to assess the GDPR compliance; – DPAs would have a structured approach to detect potential violations.”	☒	OWL	✓
PrOnto (Privacy Ontology)	“The PrOnto (Privacy Ontology) provides concepts regarding legal privacy compliance associated with data types and documents, agents and roles, processing purposes, legal bases, processing operations, and deontic operations for modelling rights and duties.”	☒	OWL	✗
Compliance Ontology/Information Model Ontology/Policy Model Ontology	“The Compliance Ontology documented in detail in the Deliverable D3.1 “Compliance ontology specification” is a generic and, at the same time, highly expressive model ultimately grounded on the analysis of the GDPR that could be easily mapped to the underlying domain-specific information model of any organisation; it actually provides a high-level codification of the GDPR, by extracting the concepts that need to be addressed by the BPR4GDPR policy framework, as well as by the privacy-aware process reengineering.”	✗	OWL	✗
Fiesta-Priv ontology	“a. Inspired by the GDPR requirements, we propose an IoT ontology built using available standards that enhances privacy, enables semantic interoperability between IoT deployments and supports the development of privacypreserving experimental IoT applications.”	✗	OWL	✓
BIoT	“...the ontology provided insight into security properties to monitor vulnerabilities in the IoT ecosystem and blockchain network structure, thereby ensuring data integrity, confidentiality, and privacy. “	✓	OWL	✗

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

5.3 SOCIAL MEDIA COPYRIGHT PROTECTION

Here, we give a literature overview for social media copyright using Blockchain technologies. Due to the small number of publications, the overview has been broadened beyond blockchain to also include decentralisation technologies to get a wider view of the topic. This review does not cover the dimension regarding the connection of ONTOCHAIN with Semantic Web technologies and ontologies for copyright management Our intention, to streamline development and build on top of recent successful experiences, is to use the Copyright Ontology and related semantic technology implementations.

The results for the Scopus query regarding social media copyright and blockchain include 3 articles and 3 conference papers. The 6 papers are listed in Table 1 below and analysed individually explained here after the tale.

TABLE 1: LITERATURE ON COPYRIGHT AND BLOCKCHAIN

Authors	Source title	Year	Document Type
Kripa M., Nidhin Mahesh A., Ramaguru R., Amritha P.P.	Blockchain Framework for Social Media DRM Based on Secret Sharing	2021	Conference Paper
Dobre R.A., Preda R.O., Badea R.A., Stanciu M., Brumaru A.	Blockchain-Based Image Copyright Protection System using JPEG Resistant Digital Signature	2020	Conference Paper
Konashevych O.	Constraints and benefits of the blockchain use for real estate and property rights	2020	Article
Daskal E., Wentrup R., Shefet D.	Taming the Internet Trolls with an Internet Ombudsperson: Ethical Social Media Regulation	2020	Article
Li Y.	The age of remix and copyright law reform	2020	Article
García R., Gil R.	Social media copyright management using semantic web and blockchain	**2019**	Conference Paper

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

The author Kripa et al. (2021) aims to facilitate copyright protection of social media content like images, videos, and audio. Copyright infringement is common in this context and protecting the content for its originality and authenticity should be simplified.

This paper proposes a blockchain framework with smart contracts to protect social media contents using IPFS, a decentralized file storage system. Content uploaded to IPFS is securely stored using a secret sharing scheme. This is combined with a robust hash for images, a method of hashing images that is resistant to modification, rotation, color alteration. The objective is to make it possible to detect near copies and block the registration of images that might be copies of previously registered ones. Unfortunately, no further information is provided about the robustness of this algorithm and its implications from a legal standpoint when infringements are not properly detected, or un-infringing content is considered otherwise.

The paper also proposes an identity management framework to generate identifiers based on a user identifier combined with a content hash and a robust hash based on perceptual features. However, there is no way to link the user identifier to a legal identity or ensure the provided user identifier corresponds to one the user controls, for instance, a social media profile.

Finally, though the use of smart contracts is mentioned as the way to ask permission to reuse registered images, no further details are provided about these smart contracts, or the way reuse terms are negotiated and then stored on-chain to provide trust to those agreements.

The author Dobre et al. (2020) proposes a system to fight intended or accidental image copyright infringement in social media platforms, mainly when professional images are used to increase the impact of posts. Photographers can use it to register their photos and users can use it to check if the image they want to use is copyright protected or not.

The main contribution of this paper is an algorithm that can extract a signature that is resistant to different levels of JPEG compression. The signature is stored on the blockchain along with the identification data of the copyright owner. It can be then used to detect copies when someone tries to register the same or a similar image, as determined by the algorithm. The algorithm can be also used by reusers to check if an image is already registered. In that case, the system allows the purchase of the right to use the photo.

The paper does not consider mechanisms to deal with situations when the registration considered a copied is the original one, giving rise to potential complaints. Moreover, the paper focuses just on the detection mechanism and little details are provided about how registrations are

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

stored on-chain, or how reuses are negotiated and managed using the smart contracts that are mentioned.

The author Konashevych et al. (2020) aims to research the possibilities of blockchain and other distributed ledger technologies (DLT) for different purposes in real estate, property rights and public registries. Though the main topic deviates from media copyright, the legal implications analysed for the real state domain are also interesting from the copyright perspective.

For instance, it is stated that the application of blockchain requires addressing digital identity and privacy issues, legal compliance, and enforceability of smart contracts. No legal enforcement is possible without solutions for digital identities and trust services, including existing regulations that already consider Public Key infrastructures. This applies to copyright too, to achieve legal enforcement in case of litigation, the authorship claims, like ownership claims for real estate, should provide evidence that links them to legal entities.

The paper also addresses the case of open blockchain and other DLTs that are open for reading. Any data becomes exposed in this case and cannot be removed. Therefore, ledgers are not suitable for storing personal data. Otherwise, the right to be forgotten (GDPR) is not applicable. The use of DLT requires some technologies and methods that are privacy preserving. For example, encrypting personal data and just storing on-chain the resulting cryptographic hash. This will provide a one-way link to the personal data, but the personal data will be stored just off-chain.

The author Daskal et al. (2020) addresses the fact that social media platforms have a powerful role as one of the main gatekeepers of the online sphere, alongside search engines and Internet service providers. Existing regulatory mechanisms, technological tools, and non-legal ethical guidelines employed in regulating content on social media platforms are not enough, there is a need for a consistent unified formal online content regulation structure that addresses increasingly threatening issues like disinformation.

As an alternative to existing regulations, tools or guidelines, the paper proposes the Internet ombudsperson as a new transnational cross-border policy mechanism that can deal with this kind of challenge. This Internet ombudsperson can be implemented following the traditional national (mainly offline) model as well as an international blockchain model.

Following a blockchain model, there should be a decentralized network composed of geographically distributed ombudspersons (nodes), who would make their assessments. Once a request is received (e.g., complaints from users, governmental requests for content removal, or a platform request for a consultation) concerning a specific ethical issue, each

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

node (ombudsperson) would present their assessment and a consensus mechanism would take place, including rewards and penalties.

The author Li et al. (2020) focuses on the emergence of remix as a dominant force of creation in the digital and Internet age. This is an issue from the copyright perspective as many scholars have agreed that the derivative works right should not cover remixes because it only protects "exact or near-exact duplication". The alternative, fair use, is a legal remedy to defend against an infringement charge, not a legal right to create something new with certainty and assurance. Consequently, fair use cannot be used as an a priori mechanisms to clear the intended remix use. For remixers, their only option is to go through a clearance process with the copyright holders of original works.

To address this situation, the article proposes that some elements of compulsory licensing and Creative Commons are combined to create a new remix rights system. This right allows the remixer to remix without permission from the copyright holder of the original work but requiring proper attribution and remuneration to the copyright holders of the reused works, like with compulsory licensing. Moreover, the remixer should allow future remixers to use the remix under the same terms and conditions received from the original copyright holder, like in some Creative Commons licenses.

The authors Garcia and Gil et al. (2019) addresses the social media copyright management using blockchain. It also stresses the fact that solutions based on distributed ledgers require sophisticated tools for data modelling and integration that can be overcome using semantic and Linked Data technologies.

In the copyright management domain, the proposal is to use the Copyright Ontology. Therefore, the idea is to build applications that benefit from both worlds, rich information modelling and reasoning together with immutable and accountable information storage that provides trust and confidence in the modelled rights statements.

The paper reports about the experience gained in this regard during the InVID H2020 European project (García et al., 2019). The results for this

5.4 ORACLES AND DECENTRALIZED ORACLE NETWORKS

In many scenarios blockchain (BC) smart contracts need to receive and process data from outside the BC network (Al-Breiki et al., 2020), (Zhang et al. et al., 2016), (Yamashita et al., 2019). However, since BC networks are isolated from the outside world smart contracts cannot obtain data from outside sources by themselves. To overcome this limitation many BC based applications use mechanisms known as BC oracles. Oracles, in the context of BC, are external data agents that collect real-world data and insert them into the BC as transactions so that smart contracts can utilize them. Examples of collected external data are gold price,

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

sensors' readings, and weather conditions. An issue that oracles introduce in every BC setting is the fact that they essentially are a third party that all participants involved in relevant transactions must trust. Below we describe the most widely known BC platforms with their respective oracle mechanisms.

Provable is an Ethereum based oracle service that provides a safe data-transport layer to enable smart contracts to retrieve external data from Web APIs. Its main objective is to ensure the availability of verifiable and auditable off-chain computation archives. Provable uses Trusted Execution Environments (TEE) and auditable virtual machines to build authenticity proofs. Even though Provable guarantees the auditable provisioning of data, limitations of current BC systems degrade Provable's performance. More specifically, issues of Ethereum such as inefficiency to handle precision-bound floating-point numbers, opcodes, the requirement to use minimal gas, high cost of operating, and lack of scalability are major bottlenecks.

TownCrier is an oracle architecture that facilitates secure data transfer between the BC and HTTP-enabled data sources such as websites (Zhang et al., 2016). The core logic of TownCrier executes as trust code in Intel's Software Guard Extensions (SGX) enclave on TownCrier's server. The SGX enclaves inherit the black box implementation properties of a TEE where neither the operating system nor other applications can interfere with the applications running inside TownCrier enclaves. The obvious shortcoming of TownCrier is that it is compatible only with settings that incorporate Intel's SGX thus excluding all others.

Microsoft Bletchley is a BC implementation by Microsoft that incorporates oracles known as Cryptlets (Gray M., 2016). Cryptlets are off-chain components that execute within a secure and trusted container and communicate through secure channels. Smart contracts and UTXO systems use Cryptlets when additional functionalities or information is needed. There are two types of cryptlets: 1) The Utility Crypltet and 2) the Contract Cryptlet. The utility Cryptlet makes up the bulk of the BC providing horizontal services like encryption, time and date events, access to external data, and authentication services. Contract Cryptlets are full delegation engines that act as Smart Contract surrogates off the chain. Additionally, contract Cryptlets provide all the execution logic and securely store the data in a Smart Contract.

The Corda BC platform has its oracles which are designed and implemented by having in mind the need to preserve privacy (Corda R3 Documentation, 2021). To this end, Corda's oracles do not have access to every part of the transaction and the only information they need to see are commands relevant to their task. Moreover, the Corda architecture guarantees that all the commands requiring a signature from the oracle in question should be visible to it but not the rest. Corda achieves that by using filtered transactions, in which the transactions proposer uses a nested Merkle

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

tree approach to isolate the parts of the transaction unrelated to the oracles task.

Gnosis (2021) is an oracle platform built on Ethereum and consequently utilizes Ether cryptocurrency. Gnosis' main field of application are prediction marketplaces. Its oracles are humans in the sense that they report information to the entity that requested them. Gnosis allows parties to dispute the reported information for a 100- ether fee. This discourages malicious disputes due to the high fee required but may prevent honest parties from disputing evidently incorrect information. Moreover, as in ChainLink's case Gnosis brings the human error in the system by giving humans the sensitive role of oracle.

Witnet (de Pedro, Levi and Cuende, 2017) implements a Decentralized Oracle Network (DON), in which a SC relates to external sources to retrieve off-chain data. Data providers are here called witnesses, and they perform Retrieve-Attest-Deliver operations, i.e., acquire off-chain data, verify it, agree on its veracity, and supply this information to the requester. Each witness is assigned a reputation value and contributes with its mining power, which in turn depends on the witness's reputation. The implemented reputation system rewards witnesses reporting correct values, while penalizes others. For each reported value, each witness weights the reported value based on the reputation of the reporter. The reputation system is exploited by Witnet to provide trustworthiness and honesty of participants, avoiding having a centralized orchestrated trust model. Witnet runs on a private blockchain network, with its own currency.

Aeternity (2017) is an open-source decentralized application (dApp) which exploits consensus algorithms to agree on the state of the off-chain world. Oracles in Aeternity are polled by SCs and retrieve data from various providers. The oracle that poses the query needs to fund its request, as well does the oracle who submits the answer. From the time that the answer has been submitted up to a certain threshold value, any other oracle can submit counter claims by investing the same amount. If no counter claims are received, then the submitted data is deemed as true, otherwise the consensus algorithm is exploited to decide on the veracity of the data.

Astraea (Adler et al., 2018) implements a DON which supports off-chain data feeds from multiple oracles. Each reported value is represented as a Boolean proposition, whereby oracles reports the outcomes of external events that can be validated in a true/false frame. Entities in Astraea are organized in submitters, voters, and certifiers. The first, i.e., submitters, allocate some stake to fund the effort of validating the submitted proposition. A small stake is also placed by voters, that play a low-risk low-reward game to vote on the proposition's truth. Certifiers instead play a high-risk high-reward game, and place a high stake in case they want to certify a proposition. However, collusion attacks may

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

harm the Nash equilibrium, resulting in false data being injected in the blockchain.

Augur (Peterson et al., 2015) implements a DON by exploiting the wisdom of the crowd. Users with reputation tokens can invest on the actual observed outcomes of Augur's prediction markets. Augur's market lifetime is split in four different phases, namely market creation, trading, reporting and settlement. Each user can, at the initial phase, create well-defined and objective events with clear outcomes. Furthermore, they can select a designated reporter. After that an ordered book of every prediction market is created, Augur's oracle determines the event outcomes to reach a final settlement. Reporters providing correct reports are awarder, whereas others are penalized. However, Augur only focuses on market outcomes, therefore being limited to this application.

ChainLink (Ellis et al., 2021) implements a DON in which the trust model is distributed between the Blockchain (on-chain data) and the ChainLink nodes (off- chain data). The main purpose of this solution is to enable communications between SC and external data sources. To this aim, two types of SCs are implemented, i.e., user SC and ChainLink SC. While the former executes on-chain application and presents requests for off-chain data, the latter provide trust and security via aggregation and reputation management. Data is pushed between SCs and web-APIs in ChainLink's trust model to ensure integrity, confidentiality, and authenticity of data for SCs. A reputation model is exploited to incentivize and penalize reporting oracles, as well as maintain fairness in the reporting system.

Kaneko et al. (2019) propose a DON, in which oracles are organized in reporters and verifiers. Both reporters and verifiers need to stake a certain amount, which is here referred to as proof of participation. If the data provided by the reporter is received in a certain threshold time value, and if the data is consistent with the outcome of the consensus, the reporter is returned its initial stake plus a revenue. The same revenue model is employed for verifiers, which are rewarded in case their claim ends to be true. A reputation system is exploited also in this case, such that higher credibility is given to entities with higher reputation.

In (Al-Breiki et al., 2019) a similar solution has been proposed. However, SCs are divided into different roles, based on their scope. In particular, reputation SCs compute and record the average reputation scores for oracles; aggregator SCs handle the communication with oracles and report the reputation scores to the reputation SC; and oracle SC hold information of deployed oracles and the data they can retrieve. Lastly, oracles are deployed to retrieve off-chain data. Data is verified based on a majority mechanism, weighted by oracles' reputation.

In (Berger et al., 2020) distributed governance is used to store oracles in a contract, together with the minimum amount of required voters.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Requests are distributed in the DON by polling randomly selected subset of oracles.

Data coming from oracles is then aggregated based on the reputation score of each reporter. The outcome is then exploited to update the reputation of the reporters.

In (Woo et al., 2020) each oracle in the DON communicates with external sources via TLS. The solution exploits the Intel SGX execution environment for key management and remote attestation. A master oracle polls the slave oracles, which in turn poll sources to retrieve off-chain data. The master oracle, upon receiving all messages from slaves, verifies their signatures, perform remote attestation in the enclave, and returns the average of the received data as outcome.

In (Liu et al., 2021) a DON is exploited together with an aggregated signature scheme. A SC polls the oracle SC to retrieve off-chain data. The oracle SC then requests data from the oracles, requiring each of them to report a signed value. The retrieved data is then checked for validity, aggregated and an aggregated signature is exploited to certify the provenance of data.

5.5 IDENTITY AND VERIFIABLE CREDENTIALS

5.5.1 Self-Sovereign Identities

Significant issues still exist to be properly solved within blockchain:

- Key management and recovery - making private key management invisible to mainstream and allowing non-custodial recovery of keys when accidents happen. Lack of good solutions in this domain has caused significant loss of funds due to accidentally lost private keys in the cryptocurrency space (Krause et al., 2021).
- Lack of accountability and identity – blockchain protocols know only about private keys (random numbers) and do not link any natural identity to accounts. This has been an intentional part of the pseudo-anonymous design of blockchains and allows censorship resistant access. However, this lack of real-world identity has caused the complete loss of accountability of humans and organizations. This, coupled with the inability to upgrade software applications on many blockchain systems has caused the irreversible loss of funds due to untraceable hacks and theft (BitNews, 2020).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

- Scalability, latency and user useability - most blockchain protocols still lack the scalability requirements to act as real currencies or to meet operation throughput requirements for their intended use case. In addition, transaction consensus and finality often take significantly longer than would be expected in web2.0 applications causing mainstream users to walk away. Lastly, and related to key management, most blockchain's do not supports functionalities that allow wallets to make the complexity of the cryptography and blockchain's invisible.

There are now a variety of protocols allowing the required scalability and key management features. There is good documentation and tooling to allow blockchain up will developers to build applications at lower cost. These blockchain protocols, combined with an SSI application architecture is the missing key to allow Gimly ID as explained in the rest of this section.

Below, we will provide an understanding of what self-sovereign identity is, how it broadly works and why it can help address the above described problems. While we provide a high-level overview of the SSI stack and architecture, it is not meant to be authoritative nor exhaustive. It is not within scope of this report to explain all of the deep details of the SSI architecture and standards. More in-depth overviews of the SSI technology stack and architecture are further provided by Tykn (2021), the Identity Management Institute (2021), and Microsoft (2021a, 2021b). For a complete understanding of the technical aspects, find the full standard specifications and architecture in the W3C specifications of DID-core (W3C, 2021) and the VC data model (W3C, 2019) and the DIF specifications of DID-Comm (DIF, 2021).

5.5.1.1 Legacy Digital Identity: The problems

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

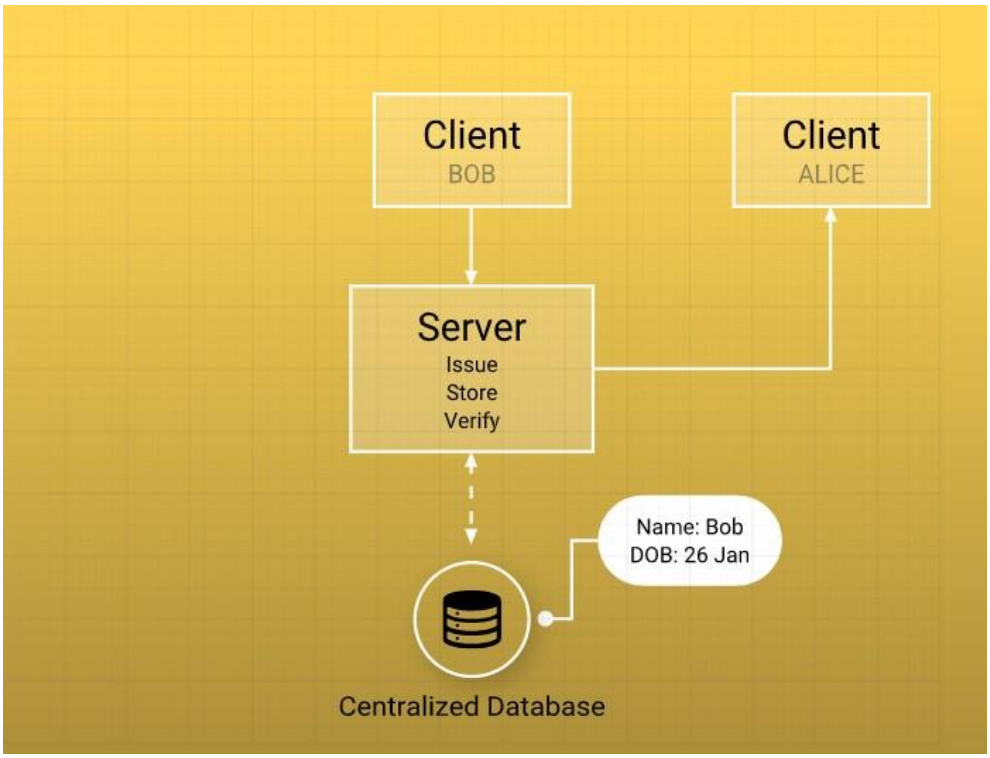

FIGURE 3: LEGACY DIGITAL IDENTITY APPLICATION ARCHITECTURE

Figure 3 illustrates the current typical software application architecture. In this architecture digital identity is stored in a centralized database operated by the service provider. When users in a system interact with another, they authorize that the service provider process their digital identity data, and check that it meets compliance, business logic and other conditions necessary for their interaction.

There are several severe interned issues with this architecture:

1. User has no control over their data, the data is centrally stored by third party and can be used by that third party for whatever purpose
2. Data processing by the service provider is opaque and the user has no control over these operations. It is also impossible to know how long organizations keep your data and for what purpose.
3. The identity data is silos and creates a single point of failure. The centralized availability of identity data has created huge honeypots and resulted in massive data breaches resulting in numerous expensive compliance lawsuits and raises major ethical concerns. This creates significant risks for both the citizen's identity data as well as new entrepreneurs and business owners to manage personal data.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

5.5.1.2 Self-sovereign identity: the solution

A new software application architecture has emerged which gives the user more sovereign control over their identity and data access. In this architecture, the user's identity data is instead held by the client software layer in the application stack.

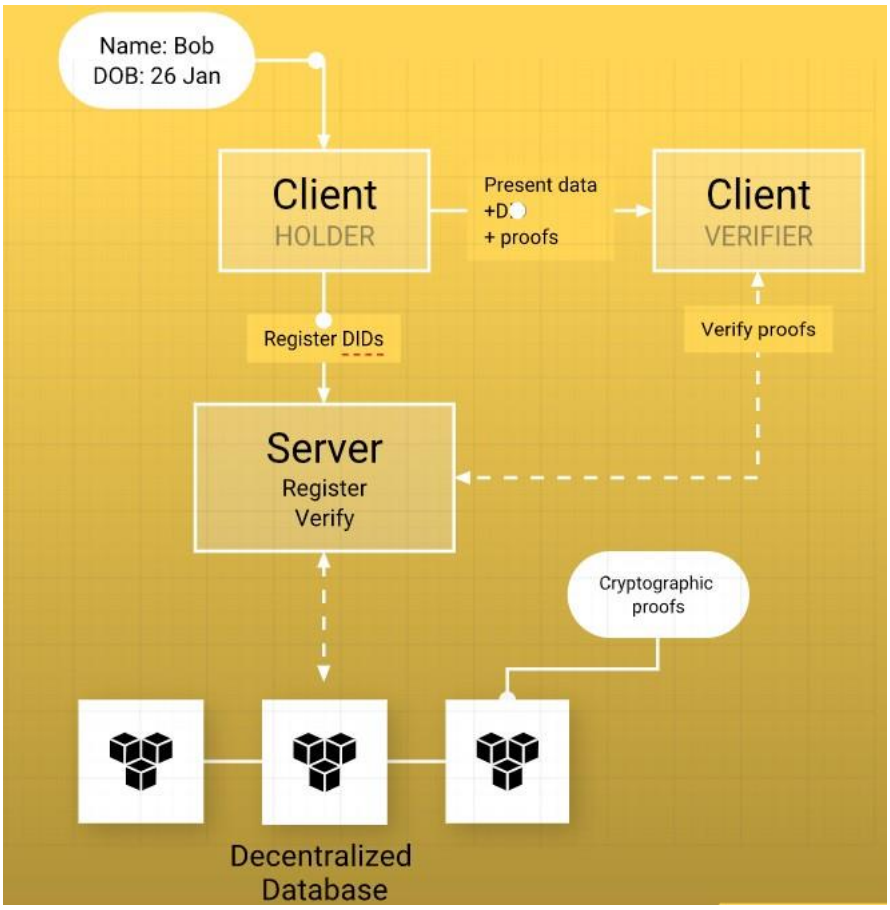

FIGURE 4: SOVEREIGN IDENTITY ALLOWS USERS CONTROL AND HOLD THEIR DATA

When one user (a 'holder' of data) in a system interacts with another (a 'verifier' wishing to receive and verify the data), the holder directly sends their identity data to the verifier who can process the data to ensure that it meets compliance, business logic and other conditions necessary for their interaction. This is a more natural way of interacting, imitating the way we interact with natural humans in the physical world. Each user can cryptographically verify the other uses data, check that it was not tampered with, without the need for the service provider. This is enabled using asymmetric cryptography and distributed ledger technology.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

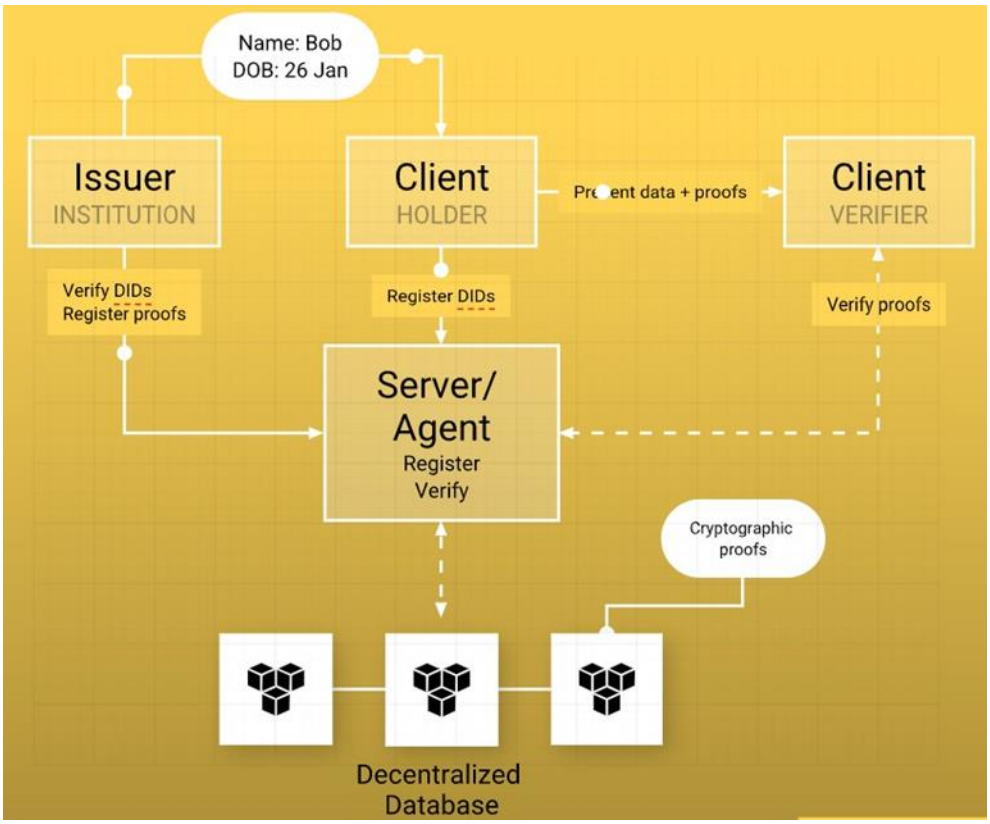

FIGURE 5: SSI NETWORK ARCHITECTURE AND ROLES

To be able to trust the data that a holder is sharing, the SSI architecture includes a third role: the issuer. The issuer can be a (public) institution providing the holder with identifying information, and registers proofs of this information which can be verified by the verifier (see Figure 5).

The SSI application stack has three layers:

1. Distributed ledger - this is the base layer and it holds the cryptographic material, built on blockchain and other distributed ledger technology. It allows users to verify sovereign data without the need for the party. It stores the public key that is corresponding to the private key which the user has in their client, and it can optionally store hashes and other cryptographic anchors as proof of identity and other data.

2. Identity - this is a way to generalise all different types of blockchain accounts into one standardized format called the Decentralized Identifier (DID). The purpose of this layer is to create a blockchain agnostic identity data structure for creating, reading and updating the public keys that control the blockchain account. This data can then be consumed by the storage

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

and exchange layer without needing to know any specific details about the decentralised public key infrastructure layer. DID formats are specified with a "DID method" which conforms to the W3C DID standard.

3. Storage and exchange – this layer allows for verifiable and blockchain agnostic identity data and exchange. There are two main W3C standardized technologies here, verifiable credentials (VC) and DIDcomm. This layer uses the DID to sign and/or encrypt data for different purposes that an identity may need such as storing information about the identity (this is done using VCs) or sending a message to in other identity (DIDComm). What is important about this layer is that a machine does not need to know the underlying blockchain technology to be able to interact between identities, and does not need to rely on blockchain infrastructure freeing such resources and reducing privacy issues.

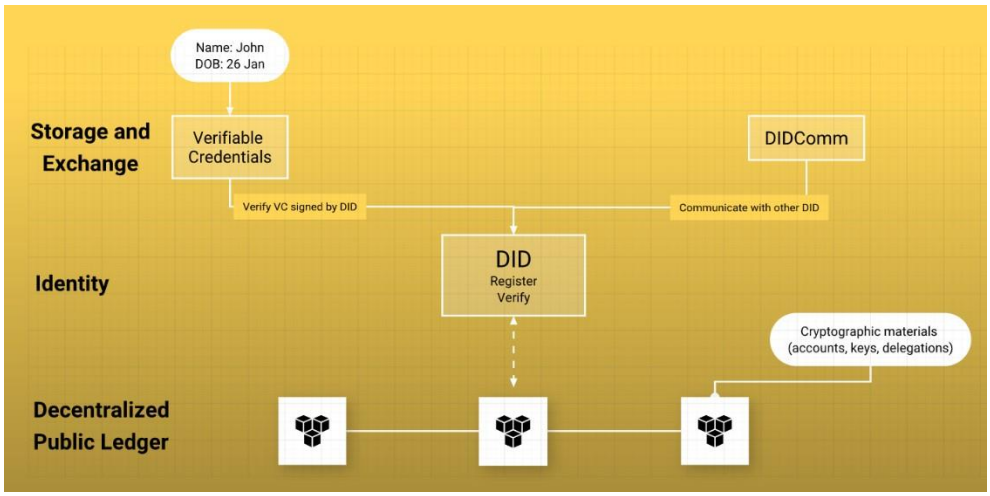

FIGURE 6: SOVEREIGN IDENTITY SOFTWARE STACK[8]

It should also be noted that SSI is not only about human identity alone, but is able to identify all relevant entities in an ecosystem, such as governments, government departments, organizations, IoT devices, natural or man-made resources or anything else can be represented by DIDs and VCs in SSI (see Figure 6).

Important to ONTOCHAIN, VC is a W3C approved standard with a rich set of existing semantic strategies. These leverage JSON-LD as the primary means of allowing machines to interpret identity and other data. Many libraries also support existing semantic ontology is such as Schema.org. With the large amount of existing worked in digital identities that has gone into

[8] Note: This is different to the network architecture diagrams of Figures 3-5.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

SSI and Verifiable Credentials, we are confident that we are building on a future proof system for representing semantic data.

SSI uses Internet standards which are collaborated on by hundreds of developers, eventually becoming a W3C approved standard. These standards have a strong focus on data privacy and security, and are reviewed by experts around the world. This standards-based approach allows digital identity is to be highly interoperable.

The SSI approach combined with blockchain has the following advantages compared to legacy digital identity systems:

1. Sovereignty - the use of asymmetric cryptography leveraging decentralised public key infrastructure, users have exclusive control over their account and authorisation. Identity data expressed in VC is under the control of the client software and is not stored on a server. As noted in section 1.3.2, sovereignty is a scale, and SSI allows for much greater sovereignty.

2. Persistence – decentralised blockchains have unique properties allowing trust in the persistence and immutability of their data. This allows digital identities to be trusted to exist without the risk that a service provider would fail or their account be censored.

3. Identity -bringing off-chain identity to anonymous blockchain accounts a sovereign way allows users to stay in control. It brings accountability back to blockchain, controlled by the users and allows for human interactions that simulate how we interact in the physical world.

4. Security and privacy - the open source and standards-based open-source approach allows researchers, developers and entrepreneurs the chance to review and scrutinise security and privacy before it becomes a problem. SSI is a premium example of this.

5. Interoperability - standards-based approach allows applications to use identities and interact between them in a way that is highly interoperable between different decentralised ledgers. The identity data itself (VCs) also becomes very easy for machines to read based on type semantic coupling and semantics technologies.

6. Scalability - taking data off the blockchain significantly reduces the processing capacity required, which thus frees it to process meaningful programs. Personal Identifiable Information should never be on a blockchain, due to its immutable nature this would at the very least violate GDPR and other existing and soon-to-be data regulation, as well as violating human privacy rights.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

SSI is relatively new, with the DID-core standard starting in 2020. Now, more than 70 DID methods have been created and SSI is starting to be appreciated and adopted at a rapid rate. It is still small, and the technology is less mature than blockchain, but the potential has been recognised by well-known enterprises like IBM and Microsoft, and pushed by governments such as the European Union, Canada, and United States of America.

5.5.1.3 SSI vs blockchain NFT example

As SSI is a fairly new approach to digital identity, we compare two different strategies, namely NFT and VC ones, for managing a digital identity that controls a blockchain account (see Table 2).

NFT strategy - create a non-fungible token (NFT) on the blockchain in in which the identity data is stored. This could be used to store attributes about identities of pets or objects, but becomes harder to store attributes about humans. NFTs are public tokens that are transacted typically, resembling a bearer token. If you own the token, you hold the attributes.

Human attributes like name, date of birth etc, are non-transferable – they always correspond to a human. The owner of an NFT has an identity through means of possession of the account/address, which by extension means possession of a private key. Using an NFT for human identity therefore has ethical consideration of whether such an NFT is meant to be a transferable token. We would strongly believe not, but still make this comparison for the sake of understanding Verifiable Credentials.

VC strategy - create a DID (a blockchain agnostic unique identifier), which possesses one or more VCs in the wallet describing the identity. These VCs contain the attributes belonging to the DID and don't live on a ledger (the proofs can). As a DID is a unique URI, pointing to a document that can contain multiple public keys, it means key recovery and rotation over time without having to change the DID becomes possible. Something which an account/address based blockchain system cannot.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

TABLE 2: COMPARISON OF NFT AND VC APPROACH TO DIGITAL IDENTITY DATA

	NFT	VC
Privacy	Very bad - identity data is public on the blockchain and immutable	Good - identity data is stored off the chain and can be verified using the public keys found on the blockchain
Scalability	Lower - blockchain infrastructure requires additional resources to process identity data	Higher - no blockchain resources are required to process identity data
Interoperability	Silod - identity data exists on one blockchain	Higher - identity data is standards based and blockchain agnostic
Semantics	Silod - semantics depend on the separate blockchain standards used	High - semantics are based on a large international community and semantic technologies
Key Management	Not part of NFT – an account's keys can be lost without recoverability of the NFTs they control	A bit easier – DIDs allow additional mechanisms for key recovery of accounts not available on blockchains
Simplicity	Much simpler – an NFT is an owned token in a smart contract on a public blockchain	More complex – an NFT is a VC object held by an identity digitally issued by its owner tied to a blockchain through DIDs

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Transferability	Easily transferable – with a one blockchain transaction	Low – possible but no standard mechanism for a VC ownership transfer exists.

5.5.2 Know-Your-Customer

Identities, or addresses, on blockchains are pseudo-anonymous and do not require their participants to get identified. However, in order to exchange fiat money into cryptocurrencies on centralized exchanges like Binance or Kraken, it is required to successfully complete a Know-YourCustomer identification.

With Blockchain Technology becoming more popular, it is necessary to take a look at how to entangle a legal identity with a blockchain address. A blockchain that supports legally binding agreements between parties is yet still not existent.

Legally verifying identities is only required if fiat currencies are being exchanged from or to cryptocurrencies on centralized exchanges or if a person does things that fall under a regulatory umbrella like profit sharing and real estate transactions. This is done by a "Know-Your-Customer" (KYC) verification where the user needs to upload documents that prove his or her identity and often the current residence. The KYC serves to prevent money laundering and terrorist financing based on the Money Laundering Act 2008[9]. Currently, these KYC requirements for blockchain-based securities management is not implemented and holds back blockchain adoption in the industry. Currently, these KYC requirements for blockchain-based securities management is not implemented and holds back blockchain adoption in the industry.

To prove the identity and to comply with KYC, the user mostly has two choices, paper-based KYC and video-based KYC. In the paper-based KYC, the user must manually upload a picture or scan of its ID document and a recent bill for his or her address. In the video-based KYC, the user performs a video-identification like IDnow[10] where the user shows the ID into a camera so that qualified service providers can look at it to verify its authenticity and confirm the information is correct. These two methods are the most common and popular ways to perform a KYC and only need to be performed once to gain access to a service.

A third way to prove a user's identity is an eID verification that could possibly be repeated multiple times. In contrast to the previous two KYC

[9] https://www.bafin.de/SharedDocs/Veroeffentlichungen/EN/Aufsichtsrecht/Gesetz/GwG_en.html
[10] https://www.idnow.io/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

variants, an eID verification can be performed automatically and can be repeated multiple times. This allows for on-demand verification.

5.5.3 Electronic Identity

An electronic Identity (eID) is meant to be the digital representation of a citizen's physical ID for interactions on the internet.

Modern IDs contain this eID in a small RFID-chip within the ID document. It can be extracted by certified services and used to automatically fill in registration-forms, to identify a user, or to prove any of a user's characteristic[11]. It bears the potential to increase data protection and the prevention of online fraud[12]. In Europe the Electronic Identification mutual recognition of eID. This allows users to use eID-based services in other EU Member States than the user's[13].

Currently, there is no productive-ready link between the use of a blockchain and its identities and with the eID. However, first service providers are aiming to implement pilot solutions such as the "IDunion"[14] consortium or "commerc.io"[15].

Government-supported eIDs are already widely accessible in Europe but barely used. The paper from Michael Kupenberg et al. "Blockchain Usage for Government-Issued Electronic IDs: A Survey"[16] is summarized which country has an eID in use and what to do with it. Table 3 below is being replicated for this use.

TABLE 3: GOVERNMENT-ISSUED EIDS

Authority that issues eID	United Arab Emirates Governm.	Estonian Governm.	Finnish Immigration Services	Government of Luxembourg via LuxTrust	Switzerland City of Zug	Switzerland Canon of Schaffhausen
Rollout status	Pilot starting 2020	In use	In use till 30.04.2019	Pilot Phase	Pilot Phase	In use
Level	National	National	National	National	Municipal	Canton

[11] https://www.personalausweisportal.de/Webs/PA/EN/eid_applications/eid_applications-node.html
[12] https://digital-strategy.ec.europa.eu/en/policies/e-identification
[13] https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eID
[14] https://idunion.org/
[15] https://commerc.io/
[16] https://www.dbsystel.de/resource/blob/5169670/47330c9be63205c5ccd5c70d288b3209/BlockchainUsage-for-Government-Issued-Electronic-IDs-A-Survey-data.pdf

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Solution used for	Digital passport "ID locker"	Data integrity, timestamp	Unique Digital Identities	Trusted blockchain identities	Self-Sovereign Identities	Self-Sovereign Identities
DLT stores identity data	No	No	No	No	No	No
DLT stores authorizations for eID	Planned	No (but the DLT acts as an access log)	No (but the DLT stores payment transaction history	No	No	Yes
eID capabilities for DLT cryptography	No	No	No	Yes	Yes	Unspecified

TABLE 4: COMPARISON TABLE OF EID-USING BL-USING BLOCKCHAIN SOLUTIONS[17]

The "German eIDAS-Middleware" implements an adjusted eID-Server with an eIDAS interface according to predefined standards and realises the server-sided component of the authentication process with the online identification functionality[18]. This process is shown in Figure 7.

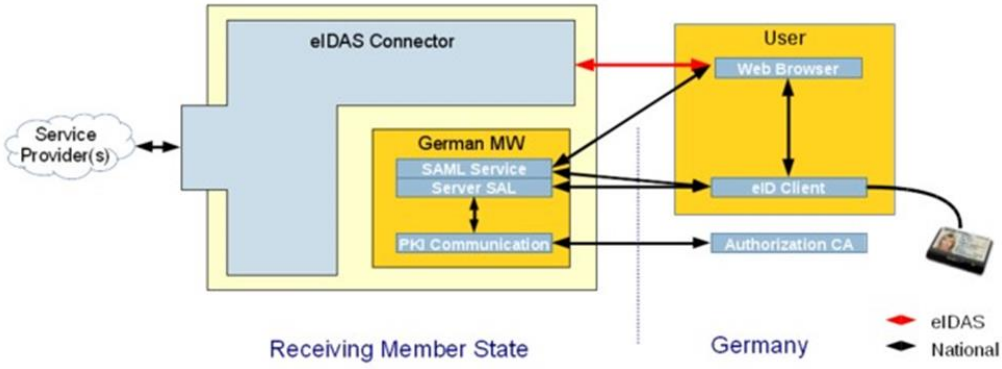

FIGURE 7: GERMAN INTEGRATION OF THE EIDAS-MIDDLEWARE IN THE EIDAS NETWORK

[17] https://www.dbsystel.de/resource/blob/5169670/47330c9be63205c5ccd5c70d288b3209/Blockchain-Usage-for-Government-Issued-Electronic-IDs-A-Survey-data.pdf
[18] https://www.bsi.bund.de/DE/Themen/OeffentlicheVerwaltung/Elektronische-Identitaeten/Online-Ausweisfunktion/eIDAS-Notifizierung/eidasnotifizierung_node.html

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Furthermore, the EU has a list of pre-notified and notified eID schemes under eIDAS[19] that could be consumed.

5.5.4 Decentralized Key Management

Key management is yet one of the unsolved problems on the internet and especially decentralized solutions. Nowadays, this problem is handled by costly password-reset mechanisms, customer-service centers or just not handled at all which could lead to total loss of funds and the identity. Conventional Backup and Recovery mechanisms can be summarized in the following criteria:

1. Physical Backup: A user writes down recovery & seed phrases and stores it somewhere save, like a vault in a bank. This leaves the risk that the recovery data gets inaccessible for various reasons like burglary, fire, or loss.
2. Cloud Services and Password Management Tools: User can store their secrets on cloud infrastructure like Google Drive, Dropbox, or similar services. Not only does the user depend on remembering the login credentials, but the user also must trust that those services don't spy on their users or will be successfully attacked, and the data gets exported.
3. Social Recovery: This process describes that a user selects a set of known and trustworthy persons or devices and provides them with parts of the secret. The secret is often shared using a threshold algorithm like Shamir Secret Sharing were each person or device receives one unique part of it. This comes close to a decentralized approach to store keys but highly depends on the trust and availability of the persons or devices that have received a share of the secret. Eventually, those parties are as prone to the risks mentioned above in addition to the trust that the contact persons don't just come together and reassemble the keys by themselves. Lastly, social relationships change over time. Who a user chose today might not be available tomorrow? There's a reliability problem when relying on people.

Since these measures leave many things that can go wrong, it is not surprisingly that Chainalysis estimated in 2018 that 3,79 Bitcoins may be lost forever[20]. Calculated with roughly 50.000€ for one Bitcoin in April 2021, this results in a loss of around 189.500.000.000€ (189 Billion €) worth of Bitcoin. And this is only Bitcoin alone, not

[19] https://ec.europa.eu/cefdigital/wiki/display/EIDCOMMUNITY/Overview+of+prenotified+and+notified+eID+schemes+under+eIDAS

[20] https://www.newsbtc.com/news/bitcoin/chainalysis-up-to/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

mentioning any other cryptocurrency like ETH. A loss of identity might have even more impact on an individual.

5.5.5 Solid and Verifiable Credentials

Identity management is an essential component of digital systems. Currently, the user is not in control of her/his identity: the so-called identity providers (IdPs) are responsible for identity attestation, management and use. They are in effect owners of the user identity.

Identity management implies knowing and managing a set of user information: age, residence, role, status, and so on. In the physical world, usually, we have pieces of paper or plastic or somethings similar that we carry around in our wallet and that prove one or more claims, where a claim is "a statement about a subject"[21] (e.g., "john is 30 years old").

A credential is "a set of one or more claims made by the same entity"[22]. Identity related information is owned by issuers (for example the state registry service) and are fragmented among various service providers.

Self-Sovereign Identity (SSI) aims to give back to the user the ownership of his/her identity giving control on its personal data and on the information provided to services. An issuer can certify specific user attributes. For example, an issuer using a Know Your Customer (KYC) process can certify the correspondence from a digital identity and the user's real identity. Any trusted party that needs to identify the user will be presented with the user-controlled portions of the identity. To accept the identity, the trusted party must have a trust relationship with the statement's issuer. Compared to most previous identity management systems where the service provider was at the heart of the identity model, SSI is user centric.

At the heart of the concept of self-sovereign identity are verifiable claims. The first clarification that is necessary for this context is between a claim and a verifiable claim. A claim is only a statement about a specific subject. Claims in a credential can refer to attributes (e.g., age, height), relationships (e.g., employer, citizenship) or entitlements (e.g., vaccination, tax exemption, university degree). Because the user owns and manages its claims, these claims must be verifiable, that is they are certified by the entity that released them. Therefore, a verifiable credential "is a tamper-evident credential that has authorship that can be cryptographically verified"[23].

In SSI, the user has the possibility to disclose only the information strictly necessary (for example, a user who has an attestation from the state registry service can prove that she/he is of age without revealing

[21] https://www.w3.org/TR/vc-data-model/#claims
[22] https://www.w3.org/TR/vc-data-model/#credentials
[23] https://www.w3.org/TR/vc-data-model/#dfn-verifiable-credentials

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

his date of birth). This is called selective disclosure and it is possible thanks to cryptographic methods like zero-knowledge proof.

W3C is working for an interoperable framework for decentralized identifiers (DIDs), and Verifiable Credentials (VCs).

In the W3C's terminology we have:

- One or more Issuers that are the source of credentials.
- A Holder that requests verifiable credentials to issuers, hold them in his/her digital wallet, and present proofs of claims when requested by verifiers using one or more credentials.
- One or more Verifiers that request proof to admit the Holder to services.

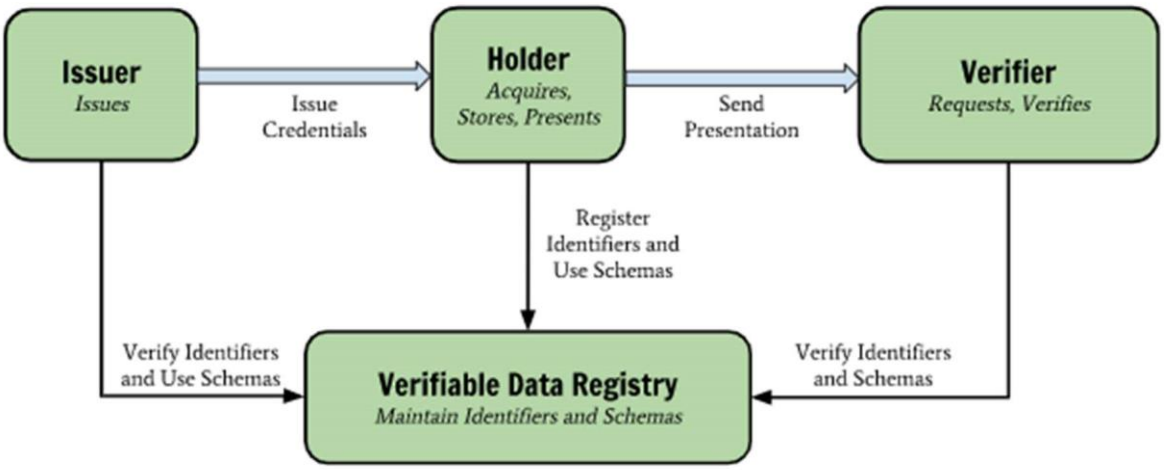

TABLE 5: VC-INVOLVED ENTITIES AND DATA FLOW (SOURCE: W3C VC DATA MODEL)

Verifiers need to identify and trust the Issuers; this is typically performed using public/private key cryptography and Public Key Infrastructures (PKIs) - the system of obtaining public key certificates from a small set of certification authorities (CAs). However, PKI is too centralized and expensive. In recent years DLTs have been proposed as a decentralized registry to hold digital identities, acting as a substitute of PKIs. We will call this function of the blockchain the register of identifiers (Identifier Registry Model). Blockchain decentralization here marries well with the aims of self-sovereign identity.

The information required to verify claims can therefore be acquired from the blockchain. Blockchain also serves as a Registry for Verifiable

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Claims. As stated, it can store public claims or only register metadata for private ones.

5.5.5.1 Overview of the W3C standard

W3C issued two standard proposals: Decentralized Identifiers (DIDs) v1.0[24] and Verifiable Credentials Data Model 1.0[25].

Decentralized identifiers (DIDs) are "identifier that enables verifiable, decentralized digital identity". In contrast to typical, federated identifiers, DIDs have been designed so that they may be decoupled from centralized registries, identity providers, and certificate authorities.

A DID is a simple text string consisting of three parts:

1) the did URI scheme identifier,
2) the identifier for the DID method, and 3) the DID method-specific identifier.

FIGURE 8: DID SCHEMA (SOURCE: W3C DID SPECIFICATION)

For example, Alastria[26], a Spanish association that promotes the digital economy through DLT, has adopted the DID model and provides identifiers like:

did:ala:fabr:testnet5:3eabc53a851fc5039eae2146083cdc42a87aeeacf848efb9924a381cc4b2b5d1

that represents an Alastria DID on its Hyperledger Fabric testnet.

Each DID is associated to a DID Document, which is a key/value pair map containing two different classes of entries:

- "A set of data describing the DID subject, including mechanisms, such as cryptographic public keys, that the DID subject or a DID

[24] https://www.w3.org/TR/did-core/
[25] https://www.w3.org/TR/vc-data-model/
[26] https://alastria.io/en/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

delegate can use to authenticate itself and prove its association with the DID"[27]

- representation-specific entries[28].

Each DID implementor can decide where to store the DID Document. Often DID Documents are stored off-chain.

In the W3C VC model, claims are expressed using subject-property-value relationships. To provide only personal data appropriate for the context at hand, the W3C VC standard envisages the so-called verifiable presentation[29] that provides a subset of personal data, even from different subject's verifiable credentials, provided in a verifiable way (see Figure 3).

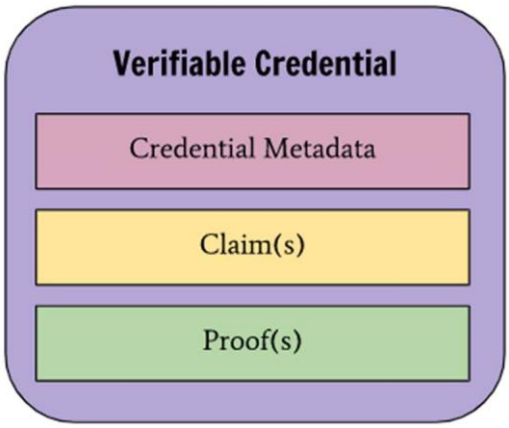

FIGURE 9: BASIC COMPONENTS OF A VERIFIABLE PRESENTATION (SOURCE: W3C VC DATA MODEL)

The following figure shows a graphical representation of a verifiable presentation stating that Pat is alumni of Example University.

[27] https://www.w3.org/TR/didcore/#dfn-did-documents
[28] https://www.w3.org/TR/didcore/#representations
[29] https://www.w3.org/TR/vc-data-model/#dfn-verifiable-presentations

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

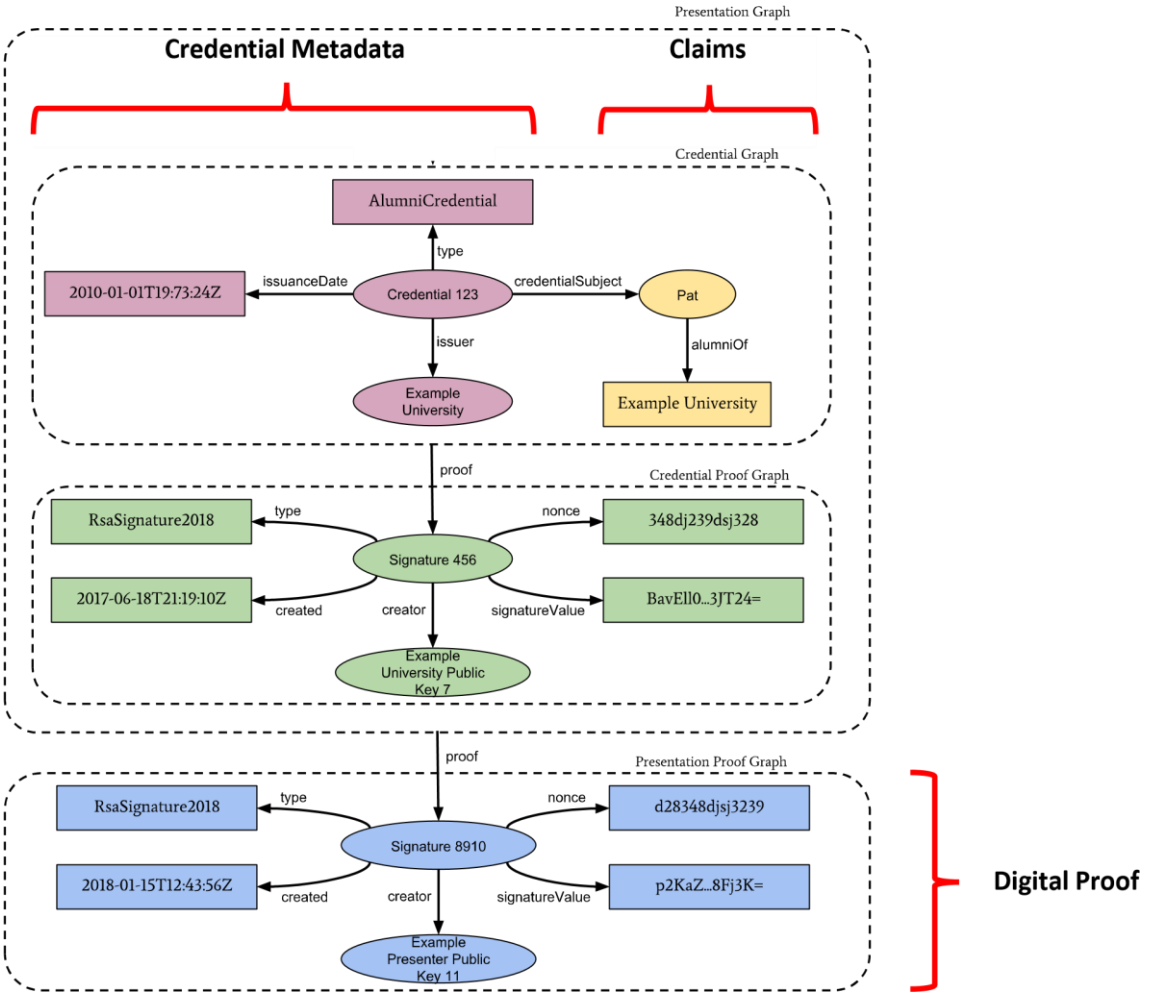

FIGURE 10: EXAMPLE OF A VERIFIABLE CREDENTIAL (SOURCE: W3C VC DATA MODEL)

The W3C standard also envisages the representation of verifiable credentials in JSON format so that they can be stored in, and managed via, a digital wallet.

5.5.5.2 Survey of present implementations

Below, we provide an overview of the most relevant SSI implementations.

SelfKey[30] developed an open-source Identity Wallet where the user can store his/her identity and the verifiable credentials that she/he receives. The wallet is a non-custodial one: user's identity documents and VCs are stored in the user device (usually a smartphone) and the user has full control over them.

[30] https://selfkey.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

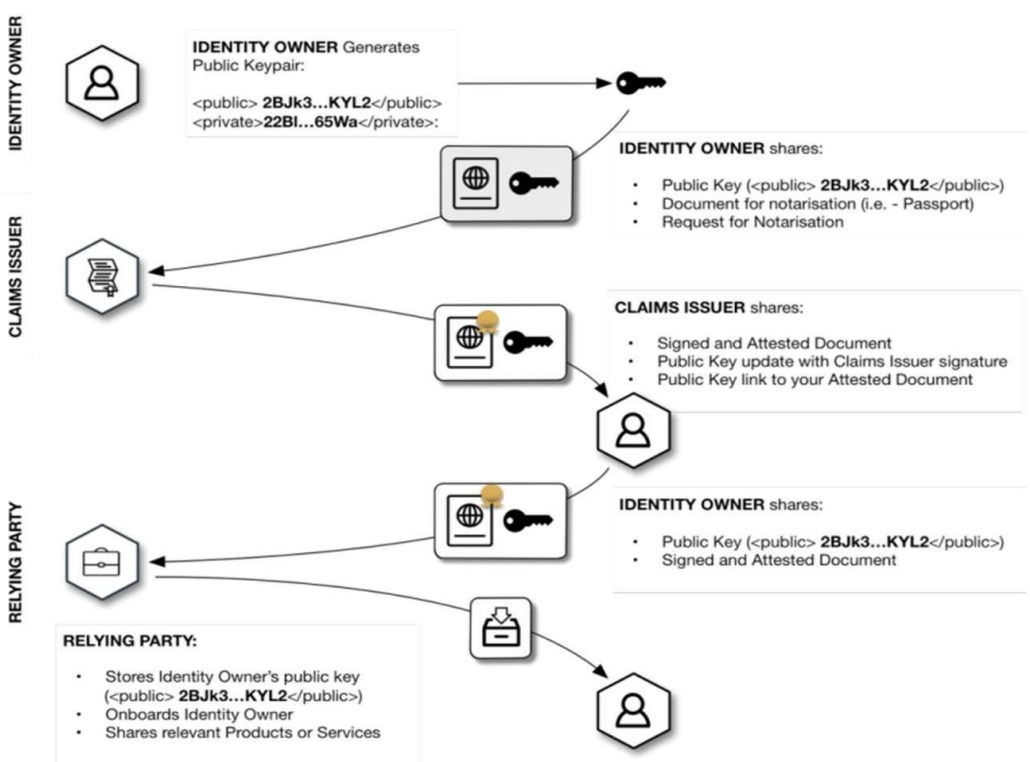

FIGURE 5: SELF KEY FLOW (source: Selfkey whitepaper[31])

Serto[32] is a well-known Ethereum based SSI. In uPort each user is represented by a proxy smart-contract that store the public key(s) of the user. So, the user global identifier is the Ethereum address (a 20byte hexadecimal string) of her/his proxy. This enables multiple keys support, key substitution and other security mechanism. Each user can have many private keys, each stored in a different personal user device (e.g., mobile phone). Information related to each identity (i.e., VC) is stored off-chain: user can choose among various off chain data storage like IPFS (a decentralized storage service), AWS (a centralized cloud solution), Dropbox and others. A Registry Contract establishes cryptographic relationship between uPort identifiers and associated off chain data attributes.

LifeID[33] is an open identity platform that implements W3C standard using a public permissionless blockchain. It stores DIDs and DID documents on chain. Also, optionally, it stores on chain hash of verifiable claims, but it does not store any user information on its network. This project does not appear currently active.

[31] https://selfkey.org/wp-content/uploads/2019/03/selfkey-whitepaper-en.pdf
[32] https://www.serto.id/
[33] https://lifeid.io/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Sovrin[34] is a public permissioned blockchain and a framework made to support SSI. It is member of the Hyperledger ecosystem. DIDs follows the W3C specification. Blockchain stores DIDs and DID documents which contain the public key for the entity and other public information the identity owner wishes to reveal to others. DID Sovrin ledger does not store any personal identity information, not even in an encrypted form. Claims definitions are stored in the ledger. Each user can have public claims and private ones. Private claims are stored in the user private portion of the ledger in a cryptographic form. Sovrin supports privacy of data by using selective disclosure for verifiable credentials which uses zero-knowledge proof.

Some of the implementation described above work on Ethereum. Looking at the standardization attempts there are some ERCs relevant for SSI. The most promising are:

- ERC-1056 - "Ethereum Lightweight Identity"[35]
- ERC-1812 – "Ethereum Verifiable Claims"[36]
- ERC-780 - "Ethereum Claims Registry"[37]

The first is a lightweight specification that proposes the use of an Ethereum address as identity. Identity creation is as simple as creating a regular key pair Ethereum account, which means that it is free (no gas costs) and all Ethereum accounts are valid identities. Furthermore, this definition is compatible with W3C DID.

An identity can have an unlimited number of delegates and attributes associated with it. A delegate is an address that is delegated for a specific time to perform some sort of function on behalf of an identity. An attribute is a piece of data associated with the identity.

ERC-1056 provides features to create and update identities via a smart contract deployed once that can be used by everyone. A user must use it if:

- she/he wants to state that an identity is owned by a different account (by default an identity is owned by itself)
- she/he wants to delegate another address to act on behalf of the user
- she/he wants associates some attributes to her/his identity
- she/he wants to revoke ownership, delegation, attributes.

[34] https://sovrin.org/
[35] https://eips.ethereum.org/EIPS/eip-1056
[36] https://eips.ethereum.org/EIPS/eip-1812
[37] https://github.com/ethereum/EIPs/issues/780

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

ERC-1812 provides an Ethereum standard for off-chain verifiable claims management. The claims are stored off-chain, but they can be verified on-chain by smart contracts.

ERC-780 provides features to allow persons, smart contracts, and machines to issue claims (including self-issued ones) via the setup of the Ethereum Claims Registry (ECR). The goal of the registry is to provide a central point of reference for on-chain claims on Ethereum. The ECR is a contract that is deployed once and can then be commonly used by everyone.

The ECR provides an interface for adding, getting, and removing claims. Essentially the registry is a key/value store where each key/value pair is registered by the issuer and it references a subject. The key parameter is used to indicate the type of claim that is being made. The proposal encourages the use of the hash of the claim type as key (e.g., keccak256('School-degree')).

There is no specification for the value that carries the claim content (value, signature and so on). It can store the claim in clear, or in a cryptographic form, it can be a signed Json Web Token (JWT), it can be a hash and a link to IPFS where there is a bigger document containing the claims, and so on.

5.6 BLOCKCHAIN AND CONFIDENTIALITY

Confidentiality is defined by the Oxford English dictionary as "the state of keeping or being kept secret or private"[38], while Wikipedia provides a more extended definition stating that "confidentiality involves a set of rules or a promise usually executed through confidentiality agreements that limits access or places restrictions on certain types of information"[39].

From the above definition it is clear that any blockchain, given its basic characteristics to make every transaction on it verifiable by any node, does not provide any embedded or implicit confidentiality apart from the one tied to the usually pseudo-anonymous user's identifiers.

Confidentiality is not equivalent to Privacy, even if often they are perceived as equivalent. To characterize both terms in a simply way we can consider that "confidentiality is about the data, and privacy is about the individual"[40]. Therefore, confidentiality pertains to a wider domain as compared to privacy (e.g., confidentiality of commercial information) (WEF, 2020).

[38] https://www.lexico.com/definition/confidentiality
[39] https://edtechmagazine.com/higher/article/2019/10/security-privacy-and-confidentiality-whats-difference
[40] https://blog.gurock.com/privacy-confidentiality-difference/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Currently the most widely approaches used to assure confidentiality in relation to blockchain are:

- Use of permissioned blockchain, where specific access control rules dictate the actions that may be taken by the users and nodes;
- Use of data hash values on the blockchain while the actual data is kept off-chain. This approach actually shifts the confidentiality measures from the blockchain to the off-chain system where access control mechanisms must be set-up;
- Use of well-established symmetric or asymmetric encryption techniques (e.g., AES, RSA). This approach can assure confidentiality even if it presents security or scalability issues. Indeed, if symmetric encryption techniques are used, then there is the need to share the keys among the involved users. If traditional asymmetric encryption techniques are used instead, multiple encryptions of the data, using different public keys, are required if the confidential data must be accessed by many users;
- (fully) homomorphic encryption[41] or secure multi-party computation (Chaum, Crépeau and Damgard, 1988) that make possible to perform confidential and distributed computations in zero-trust environments. In connection with blockchain these techniques are usually active in specific nodes that process transactions blindly, so that nodes can verify the transactions' correctness without viewing their content (Lesavre, Varin and Yaga, 2021). These techniques usually make use of Trusted Computing resources (e.g., TPM) and are quite slow (WEF, 2020).

Other relevant solutions are based on Zero-Knowledge Succint Noninteractive Argument of Knowledge (ZK-SNARKs) (Petkus, 2019) (or the ZK-STARTs alternatives) a new form of zero-knowledge cryptography that enables to generate a proof of possession of a specific information without revealing it and without any interaction required between the prover and the verifier. This means, that only one message is required from the prover to the verifier, together with public parameters. The "succint" property means that the proof is only a few hundred bytes long and that can be verified in a very short time making it suitable for blockchain applications. Zcash[42] makes use of the ZK-SNARK technique to support private transactions whose correctness can be proved without revealing any private information.

[41] https://homomorphicencryption.org/introduction/
[42] https://z.cash/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Anyway, these techniques can be used to avoid disclosing private/personal information, but are not suitable to assure controlled access to the information.

Recently, in the context of the OASIS standardization organization[43], a set of companies are working on the definition of the Baseline Protocol[44] that aims at defining a common frame of reference to support business processes across multiple companies by exchanging data among their IT systems via DLT in a confidential way. The Baseline Protocol fosters the adoption of public Mainnet, even if it aims also at be operational across different blockchains. The Baseline Protocol is still in a very early stage (still at V0.1.0[45]) and, as stated, it focuses on supporting interconnection of corporate IT systems across the blockchain.

5.7 CP-ABE ENCRYPTION MECHANISMS AND ITS STANDARDIZATION

Traditionally, access control (Sandhu and Samarati, 1994) is enforced by storing data on a trusted server that checks preventively that users present proper credentials that are verified against access rules (e.g., Role-Based Access Control RBAC (Sandhu et al., 1996)) before allowing them to access to data. However, in distributed scenarios this approach become more and more difficult to manage (Bethencourt, Sahai and Waters, 2007). To this end, Attribute Based Encryption (ABE) can be useful in providing with one single technique both confidentiality and access control.

ABE is a new, asymmetric encryption technique in which users' keys, access policies and ciphertexts are tied together via a set of attributes. These encryption algorithms make possible to encrypt an information in such a way that different decryption keys can decrypt it, and, at the same time, the decryption keys can be generated based on a set of public data elements and of a varying set of attributes.

There are two types of ABE schemes: Ciphertext Policy Attribute Based Encryption (CP-ABE), and Key-Policy Attribute-Based Encryption (KP-ABE).

CP-ABE schemes encrypt data according to an access control policy based on a list of attributes, and user's keys are associated to user's descriptive attributes. KP-ABE schemes, instead, have a reversed approach where the ciphertext is associated to sets of descriptive attributes, and users' keys are associated with policies instead.

In SEIP, CP-ABE is more appropriate, as it enables the data owner to freely define the access control policies while requiring less administrative effort as compared to KP-ABE where access policies are, instead, embedded in user's keys. In CP-ABE a user can decrypt a

[43] https://www.oasis-open.org/org/
[44] https://docs.baseline-protocol.org/
[45] https://github.com/ethereum-oasis/baseline

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

ciphertext if and only if his/her secret key matches the conditions stated in the access policy used to encrypt the data.

CP-ABE schemes mainly consist of four functions:

- {PK,MK}=Setup(): which outputs the public parameters PK and a master key MK.
- SK=Keygen(MK,S): which generates a private key SK for a given attribute list S.
- C=Encrypt(PK,M,A): which encrypts the message M and produces a ciphertext C such that only a user that has a set of attributes that satisfies the policy A will be able to decrypt C.
- M=Decrypt(PK,C,SK): which decrypts the ciphertext C and produces the plaintext message M, provided that the user's private key satisfies the access policy encoded within the ciphertext.

Usually, the attributes list associated to a private key is directly derived from the user/entity identity and profile.

For example:

Country=IT, Organization="Fincons SpA", Unit=Operators, CommonName="John Smith"

may be the attribute list associated to the user named John Smith.

The policy, instead, is a logical expression of attributes. For example, the policy:

Country=IT Organization="Fincons SpA" Unit:Operators

(CommonName="John Smith" OR CommonName="John Doe")

allows the decryption of the ciphertext by the users John Smith and John Doe of the organizational unit Operators from the Italian division of FINCONS SpA.

CP-ABE could require heavy computation, which may not be suitable for some scenarios. However, by combining CP-ABE with Advanced Encryption Standard (AES) symmetric encryption it is possible to drastically reduce encryption and decryption times (Pérez et al., 2018). Therefore, using an ephemeral key and AES to actually encrypt the data, and encrypting with CP-ABE only the ephemeral symmetric key (Figure 11) the computation and encryption/decryption time can be heavily reduced, even if the CP-ABE access control features are fully available.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

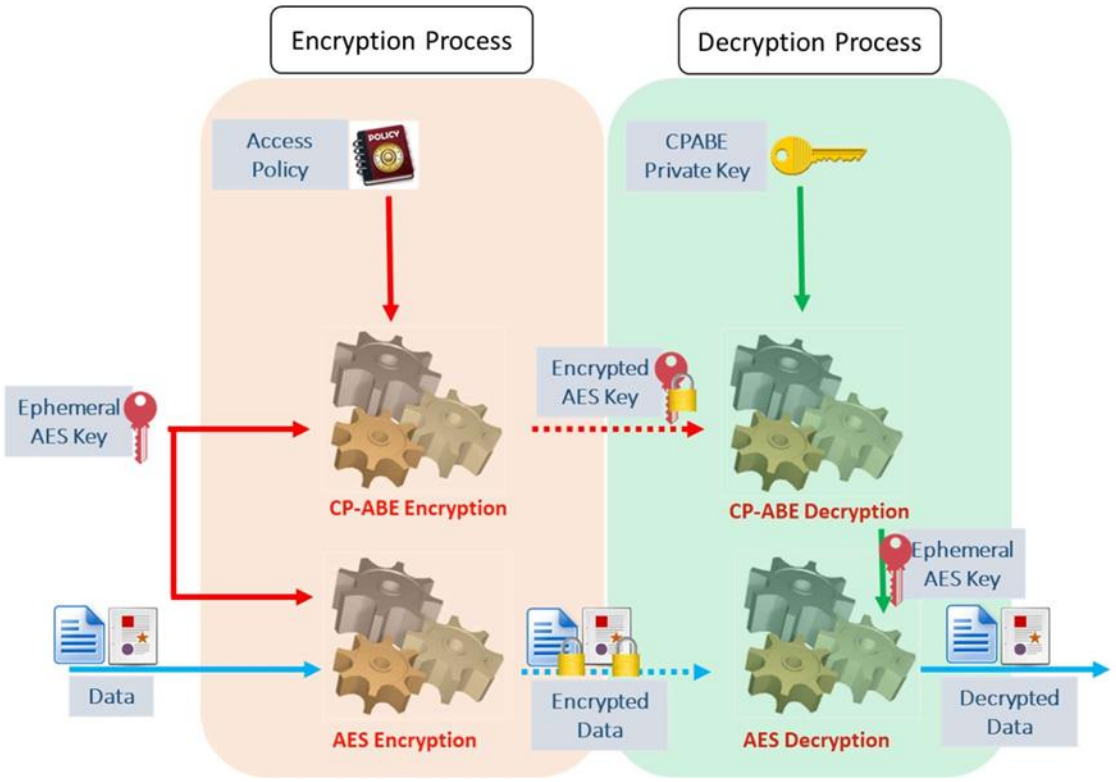

FIGURE 11: COMBINED USE OF CP-ABE AND AES

New networking paradigms are emerging such as Information-Centric Networking (ICN) (IRTF RFC7476, 2015), which are designing new Internet architectures that better support applicative scenarios such as social networking, real time audio and video communication, AR, etc. and address issues as scalability and reliability.

Information confidentiality, integrity and originality issues are, of course, present even in ICN. In this context, for example employing symmetric encryption as many disadvantages like for the example the need to share secret keys, with asymmetric encryption instead, there is a problem of redundancy of encrypted data (the same data is encrypted multiple time, one for each user). To this end, one proposed solution is indeed ABE to enforce access control to heterogeneous groups without the need of requiring multiple encryptions of the same data (IRFT RFC7954, 2016), (Ion, Zhang and Schooler, 2013), or to address information authenticity and integrity (Ramani et al., 2019). Moreover, ABE encryption, thanks to its policy mechanism, allows to encrypt data without knowing a-priori the exact entities that will be able to decrypt them.

ABE techniques are currently heavily investigated due to its features, flexibility and relations with other mechanisms based on pairing groups

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

cryptography[46]. Anyway, some standardization activity has been started. To this end it is worthwhile to mention:

- IETF activity on the use of pairing group cryptography (Sakemi et al., 2020);
- ETSI related standardization activity[47] and related standards (ETSI TS 103 532 V1.1.1, 2018) (ETSI TS 103 458 V1.1.1, 2018).

5.8 SERVICE LEVEL AGREEMENT MANAGEMENT WITH BLOCKCHAIN

An SLA documents Quality of Service (QoS) requirements as well as obligations for the parties involved in the agreement that should trigger compensation mechanisms in case of violations. The essential elements of an SLA include service type, expected service quality, conditions of service quality (peak, off-peak, maintenance time windows), metrics to measure service quality, acceptable metric benchmarks, frequency of monitoring, and penalties for breach of service quality (Ranchal et al., 2020).

Besides service quality metrics, an SLA can also include specifications for security, privacy, compliance, data backup, and disaster recovery. The SLA is composed of a 5 steps lifecycle such as: 1) SLA Definition that establishes the contract with the QoS parameters on the base of the service discovered; 2) SLA acceptance and service deployment, 3) SLA Monitoring that periodically measures the SLA metrics; 4) SLA Violation Detection that identifies any deviation from the acceptable metric benchmarks and Enforcement that takes necessary actions for violations per the contractual policies, 5) Termination that identify the conclusion of the service usage and the related SLA monitoring.

[46] https://en.wikipedia.org/wiki/Pairing-based_cryptography
[47] https://www.etsi.org/newsroom/press-releases/1328-2018-08-press-etsi-releases-cryptographicstandards-for-secure-access-control

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

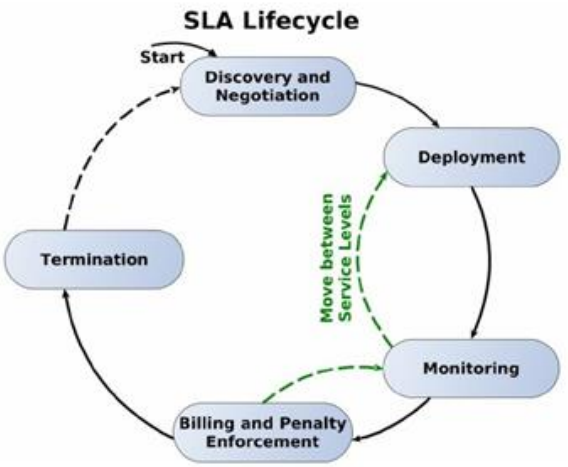

FIGURE 12: SLA LIFECYCLE

Up to now the SLA lifecycle is mostly managed through legal documents and requires manual operations for orchestration. These documents are complex, open to interpretation, and not readily enforceable. There is no automated monitoring of SLAs, and therefore SLA violations may go undetected until an adverse event occurs.

Moreover, this process requires trust between the involved parties. On the one hand, the client must provide evidence (i.e., data) that the SP is not delivering what was agreed, and thus have to allocate resources for monitoring the contracted service. On the other hand, the SP must also monitor its services to provide evidence that they are being properly provided. Therefore, reducing the number of third-parties in SLA compensation process, while providing trust in the data to involved parties, is important to reduce costs and minimize unnecessary resource allocation (Scheid et al., 2018).

However, the automation of SLA management brings a number of requirements. For instance, the individual SLAs between clients and services can be confidential and therefore not freely shared with third parties for SLA monitoring and violation detection. To this end, a centralized third-party solution for end-to-end monitoring of SLAs in a multi-cloud environment seems to be not feasible due to a lack of mutual trust and competing service provider interests. Moreover, the SLA violation should be claimed by the customers on the basis of its evidence that may lead to.

This is the scenario in which blockchain and Smart Contract (SC) come into play. In literature there are some proposals to automate SLA management using Blockchain. For instance, (Nakashima et al., 2017) implement a Smart SLA Platform (SSLAP) as a common SLA contract platform based on SC (see Figure 13). Because of the limitation of the BC to manage data, the SC refers only to the URI of the SLA to identify the

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

document on the WEB (Resource identified by URI). Payments are based on inner currency of the SC (e.g., Ether). The SLA publisher is identified through a signature.

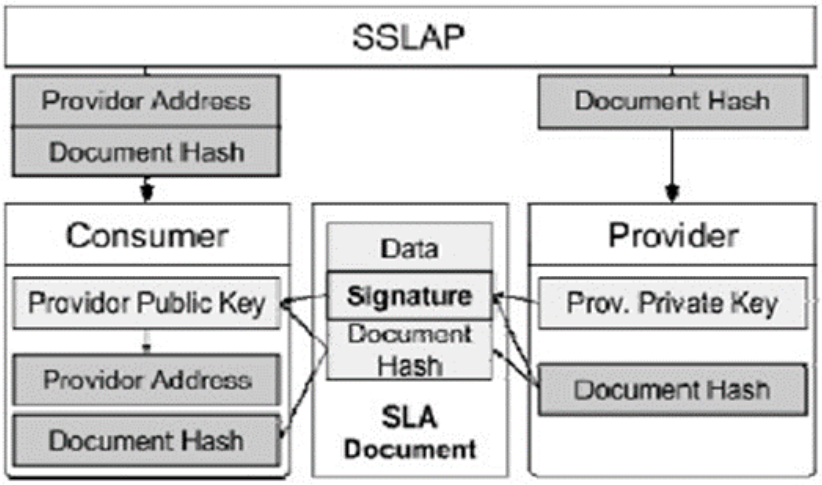

FIGURE 13: SMART SLA PLATFORM (SSLAP)

A provider gets a hash on an SLA document, and registers it to the SSLAP. SSLAP registers the publisher address and hash of the SLA document. The provider generates the signature with the hash and its own private key, and attaches it to the published document. A consumer gets the address of a provider, a hash of the document and the URI of the SLA and get the SLA document. Unfortunately, such an off-chain management introduces a risk of loss of data since it delegates the preservation and the persistence of the access to the document to a web server whose weak persistence is well known.

Kochovski et al. (2020) use the Smart Oracle (Chainlink) to collect data from the Monitoring service and to provide data to a decision making layer. However, the privacy requirement seems to be not addressed.

Ranchal et al. (2020) addressed the multicloud scenario where monitoring service in a multi-level and multi provided system is difficult due to decentralized control and lack of visibility in the entire system hierarchy. SLA violations often go undetected as multicloud systems lack automated means to monitor SLAs and conduct root cause analysis of violations. Moreover, the service providers do not freely disclose their composition and may dynamically update their component services. Therefore, the complete hierarchy of a multicloud system is not known a priori. The solution proposed is based on Hyperledger and preserves the confidentiality of SLAs (using the channel feature), yet it is able to determine the root cause of SLA violations without even having a complete view of the system topology managing the correlation of the timing on SLA violation in the hierarchy.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Uriarte et al. (2018) proposed a framework to manage dynamic SLAs in a distributed manner by relying on smart contracts and the blockchain. In particular it supports two types of blockchain environments: (a) open cloud markets where anyone can sell computational resources; (b) private deployments that comprise a single provider and several clients, some who are willing, for a financial compensation, to play the roles of auditor. In these environments, our framework employs smart contracts to automate SLA coordination.

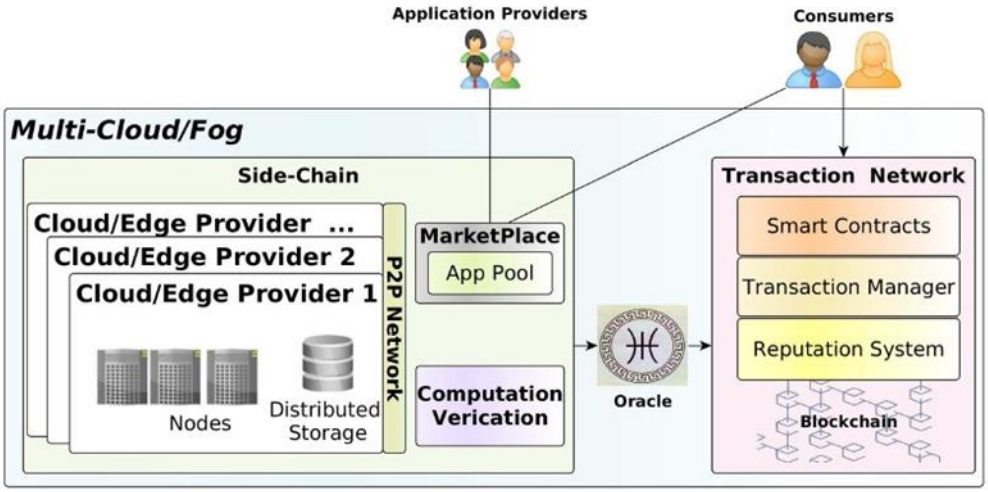

FIGURE 14: SLA CLOUD FRAMEWORK

The specification, negotiation and static verification of the SLA are executed in the side-chain via the SLA Cloud framework (see Figure 14) since these processes might require high processing power, which is particularly costly in the blockchain. For example, the matchmaking and negotiation of offers and requests may need to consider a large number of states due to the dynamic needs of the parties captured, as well as verify the compatibility and negotiate with many different providers since many offers/requests might be available. After the SLA is defined and verified, it is transformed into a smart contract, which is then executed, enforced and billed in the blockchain.

In (Zhou et al., 2018) a witness model is proposed to tackle the challenge of detecting SLA violations in a trustworthy way. A new role termed as witness is added in the traditional cloud service delivering scenario to perform as the performance monitor. The witness is designed as an anonymous participant in the system, who desires to gain revenue through offering the violation reporting service. The payoff function for different actions in our agreement model is carefully designed that the witness would have to always behave honestly in order to gain the maximum profit for himself, which can be proved by game theory.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Wonjiga et al. (2019) consider an SLA offered by the provider that guarantees the integrity of tenants' data, and propose to verify the SLA using an integrity checking method based on a distributed ledger. The proposed method allows both Cloud Service Providers and tenants to perform integrity checking without one party relying on the other. The method uses a blockchain as a distributed ledger to store evidence of data integrity. Assuming the ledger as a secure, trusted source of information, the evidence can be used to resolve conflicts between providers and tenants. They consider SLAs addressing the integrity property of data. Currently, most SLAs only address availability aspects. For example, Amazon S3 claims to have "extremely durable" storage with redundancy and checks for corruption while data is at rest and in the network. However, Amazon SLA does not guarantee this property. Hence, we assume the definition of SLAs with an objective to keep the data uncorrupted for the SLA lifetime. Other properties like backup frequency and type of access control policies can be included in the SLA definition.

In (Alzubaidi et al., 2020), Smart contracts are not considered optimal for conducting endless activities such as monitoring. Thus, external monitoring/reporting means have to be in place to help smart contracts in forming a decision on the compliance level of obligated providers. For illustration, consider the GCP SLA definition of MQTT error rate. Monitoring agents should collect metrics related to the performance of the MQTT server such as up-time, the number of received/sent messages. These metrics help the smart contract to form a decision on the compliance status of the obligated provider. That is, monitoring agents must submit any identified incidents to the smart contract. The solution proposed is based on Hyperledger.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

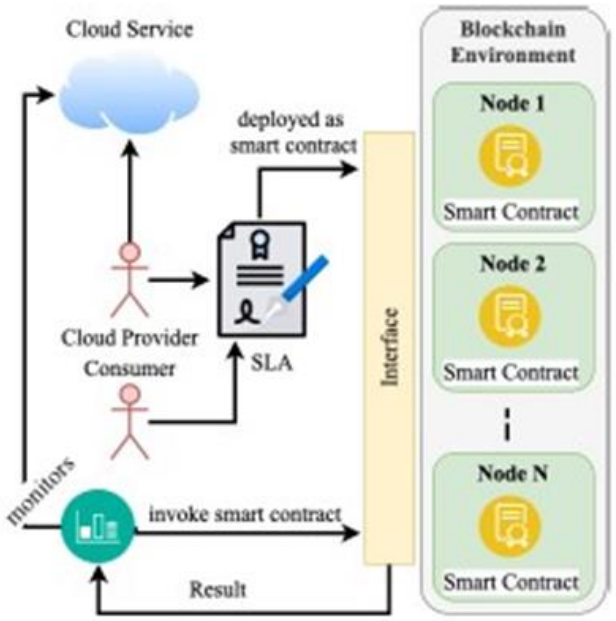

FIGURE 15: SLA COMPLIANCE CHECKING THROUGH SMART CONTRACTS

The authors (Hang et al., 2019) propose an enhanced decentralized sharing economy service using the service level agreement (SLA), which documents the services the provider will furnish and defines the service standards the provider is obligated to meet. The SLA specifications are defined as the smart contract, which facilitates multi-user collaboration and automates the process with no involvement of the third party. The solution is implemented with Hyperledger Fabric technology. We propose the use of the on-chain data lake to take away the current values of the ledger states from the blockchain. Clients submit transactions that capture changes to the on-chain data lake, and these transactions end up being committed to the blockchain. The most remarkable difference between the blockchain data structure and the on-chain data lake is the data immutability. The data cannot be modified in the blockchain even by the network admin once it is written into the ledger.

However, the data stored in the on-chain data lake updates incessantly whenever the state value changes, such as when the ownership of a car is transferred from one to another. The on-chain data lake serves as a database, which provides a rich set of operators for the storage and retrieval of states. The blockchain network can be configured to use different databases to address the needs of different types of values and the access patterns required by applications. The proposed approach aims to greatly enhance the transaction processing performance since they do not need to traverse the overall transaction log in the blockchain.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

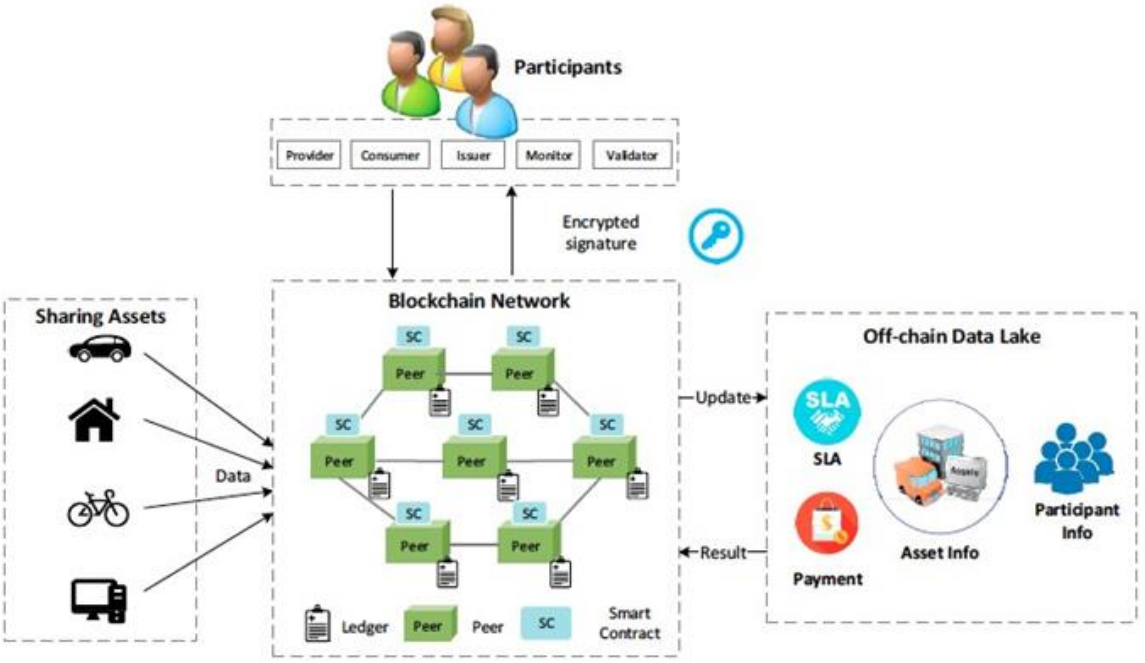

FIGURE 16: SLA COMPLIANCE CHECKING WITH OFF-CHAIN DATA

In summary:

- The issue related to data management for SLA (dimension, computational work to process) on a blockchain like Ethereum is a remarkable shortcoming. However it is addressed by adopting side-chain or off-chain on one side, or adopting Hyperledger.
- It is necessary to find the "chain of responsibility " in case of SLA failure, in particular for composed/clustered services.
- Smart Contract should not be used for computational intensive activities as monitoring but this activity is preferably delegated to Smart Oracle.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

6. BLOCKCHAIN APPLICATIONS

ONTOCHAIN architecture is a block diagram that includes platform functions, use cases and applications. Use cases are very important as they serve as a bridge between platform functions and applications.

We defined use cases to facilitate the design of the ecosystem, to establish the implementation of robust applications by orchestrating suitable certified use cases. The ONTOCHAIN framework covers various use cases related to Persons, Identity Management, Reputation Assessment (for sellers, for buyers), Data, Copyright Management, Metadata and document assessment (including the identification of fake information).

ONTOCHAIN application are based on Blockchain technology that is characterized features such as decentralization, supports disintermediation data immutability and time stamp supports the federation of data, whilst retaining ownership. Furthermore, Blockchain is the appropriate technology to unleash the possibilities of ontologies and linked data, which can support queries on data bases of different origin support reasoning. As a result of these features Blockchain technology offers opportunities to enable a new wave of applications in the Next Generation Internet (NGI) like disruptive, user-centric, to support the inclusion and the collective intelligence, data-centric, to support the data economy: in ONTOCHAIN we include ontologies and linked data not only to unleash the potentialities of these technologies, but to properly remunerate data providers (this is a weak business point today!)and more efficient, reliable and secure.

Even Blockchain technology offers a secure and trusted environment Blockchain applications face different challenges such as Scalability, Energy consumption, Security. The novelty of the technology implies also issues in the identification of new disruptive applications, the identification of most suitable business models and in the consolidation of regulations.

Applications can be categorized according to two dimensions:

1. The domain (e.g., finance, insurance, health and so) many domains can be further distinguished in subdomains.
2. The main stakeholder: Citizens, Public administration, and Enterprises

This taxonomy was used with the intention to facilitate the identification of new possible application in the scope of ONTOCHAIN objectives.
Table 6 gives a short overview on 12 domains.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

TABLE 6: VERTICALS THAT BENEFIT FROM BLOCKCHAIN

Domain	Subdomain	Application/Reasons why
Finance	Banking, Insurance	More efficient and secure operations, best customers support also by enabling new providers (Open Banking)
	Philanthropy, Social inclusion, Humanitarian support, Crowdfunding	Thrusted and decentralised solutions , easier to activate solutions
Real estate and property	Real estate	Decentralized and trusted identification and management of property rights and transfers
	Artistic assets	
	Jewelry	
Health	Medical records management, drug prescription management, clinical trials support	To leverage patients generated data in a trusted, decentralized way, with suitable privacy support
	Pandemic and chronical diseases centralized monitoring	To obtain more accurate and timely information, from properly anonymized data. Early warning can imply the use of AI
Public services	Certificates management (passports, driver's licenses, diplomas, tax certificates)	To simplify interactions of citizens with public administration. To facilitate the federation of data available by different administration bodies and nations, whilst leaving the full control of its data by any public owner.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Mobility	Safety	Efficient traffic emergencies handling (IOT and AI related)
	Efficiency	Optimization of traffic control (IOT and AI related)
Green	Energy	Electric mobility support. Decentralised energy trading, green certificates trading. Metering, billing and security (IOT related)
	Waste control	Circular economy support. Dangerous waste control.
Industry 4.0		Productivity enhancement (also with AI and IOT. Early fault detection (also with AI and IOT)
Agri food/Logistic		Product quality and tracking. Product components quality and tracking
News /Media / Entertainment	Journalism / editors Journalism / readers	Discovering fake news and biased analyses, identify original sources
	Citizens Journalism	Disintermediate citizen journalists and their public. Qualify news and journalists, discovering fake news or biased analyses.
	Media	Disintermediate performers and their public. Allow fairer performers remuneration. Facilitate the identification of new performers. Support the 'long tail'.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

	Certificated education 'How to' courses	Validate tutor capabilities and course, facilitate retrieval

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

7. BLOCKCHAIN STANDARDS

As mentioned in (WEF, 2020), blockchain standardization activities are carried out by different kind of organizations, which can be distinguished into these categories:

1) International standard organizations (ISO, ITU, IEEE, W3C, IETF) or regional standard organizations (as CEN, CENELEC, and ETSI in Europe), which, in the framework of a broader scope, develop also activities on blockchain standards
2) Industry specific bodies, which develop also some industry specific standards, as in healthcare (HIMSS) https://www.himss.org supply chain and logistic https://www.gs1.org/, and in this framework have also established an activity on blockchain
3) Specific groups or alliances in blockchain, as EEA (Enterprise Ethereum Alliance) https://entethalliance.org/ IWA (Interwork Alliance) and DIF (Digital Identity foundation) https://identity.foundation, which have also established standardization activities [48]
4) Decentralized proposals, emerging by the developer community and often facilitated by the open-source platform GitHub. Some examples include
   a. Bitcoin Improvement Proposals (BIPs) https://github.com/bitcoin/bips ,
   b. Ethereum Improvement Proposals (EIPs) https://github.com/ethereum/EIPs [49]
   c. Cash Improvement Proposals (ZIPs) https://github.com/zcash/zips

Also, the standardization activities are differentiated in different categories as:

1) Terminology and taxonomy
2) Definition of the framework architecture
3) Definition of specific blocks and functions
4) Definition of protocols and interfaces
5) Governance
6) Use Cases
7) Other

ONTOCHAIN is active in the various standardization activities around blockchain, as depicted in Table 7. ONTOCHAIN this year has attended the roundtables organized by the European Commission to familiarize with the

[48] They are summarized in the table 3 of the report of the WEF

[49] Well known examples of EIP standards are ERC-20 and ERC-721, which define APIs respectively for tokens and for non fungible tokens

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

activities of Horizon participants in different blockchain standardization activities of Standard Organizations.

Each roundtable, of typical duration of three hours, was focused on a specific vertical, e.g., e-Health, and was structured in three sections:

a) A section presenting main standardization activities in the vertical by Standard Organizations, as ISO, IEEE, CEN/CENELEC and so on.

b) A section presenting the EC initiatives to support activities for new standards, which can typically emerge towards the conclusion of a research project.

c) A section in which some Horizon-funded projects, typically towards the end of the roundtable, identified which of their results could be exploited as input for standardization activities.

The ONTOCHAIN project has considered its participation to these web conferences as an input for the status of the art analysis, but in the second and third year of the project, it will also investigate whether some of the ONTOCHAIN results could be of interest as input for standardization activities.

TABLE 7: ONTOCHAIN PARTICIPATION TO BLOCKCHAIN STANDARDIZATION ACTIVITIES

Project	Fintech, Digital Assets and Smart Grids	Digital Society, Identity and Privacy	Digital Economy, SME's, Industry and Supply chains	Cybersecurity	IoT	eHealth	Future Internet, Media and Big data
	11-Nov-20	25-Nov-20	9 Dec 2020	13-Jan-21	27-Jan-21	10 Feb 2021	24 Feb 2021
ONTOCHAIN	X	X	X	X	X	X	X

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

8. CHALLENGES

The most important challenges to realize the ONTOCHAIN framework are the following (Deshpande et al., 2017):

Interoperability:

One of the most difficult and essential jobs of semantic web ontology engineering is the integration of ontologies with the objective of generating a single ontology for all web providers and consumers in a domain. The accessible ontologies frequently have different conceptualizations of similar or overlapping topics, which creates an interoperability issue. This issue could be solved by using ontology matching which detects semantic relations between concepts, characteristics, or in-stances of two ontologies. The use of reasoning languages (e.g., Distributed Description Logics) to reason about ontology alignments in distributed environments is a current and future trend in reducing schema heterogeneity.

Scalability of semantic web applications:

Scalability is one of the most important challenges in large ontology generation and maintenance, semantic metadata extraction from enormous and diverse information, and inference techniques. The issue of scalability was discovered early in the semantic web study and was appropriately addressed. Despite the large number of semantic web apps available today, complex Semantic Web technologies like reasoning under open-world assumptions are difficult to implement real-time on the web. The logical method has dominated the semantic web until now but adopting the Information Retrieval (IR) approach might help with scalability. Other aspects that must be taken related to the scalability issue is what parameters should be monitored and what benchmark should be varied? Recently there is no agreement on which standards should be used to define the problem of scalability on the semantic web.

Large-scale adoption:

The issue of widespread semantic web technology adoption in circumstances where existing Web Technologies have already shown to be beneficial. Additional funding is required to advance semantic web technologies, particularly to improve querying and reasoning techniques. Instead of competing, the Semantic Web should incorporate elements from other related research groups, such as the Social Web and the Pragmatic Web (http://www.pragmaticweb.info/). While the Semantic Web encourages" cooperation through common models of knowledge," the Pragmatic Web relies on" shared ways of socially developing the ways of representing knowledge". As a result, future research and development should focus on

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

ontology negotiations, contextual ontologies integration, and establishing pragmatic patterns for describing concerns like communication, information, and tasks.

Distribution vs. Centralization:

Distribution is generally considered as a possible route to scalability and very well interconnected to the Web's dispersed character. All the major search engines operate by locally storing the whole Web and running indexes, etc., on that local, centralized cache. The distribution of the semantic web is an effective way to solve this issue.

9. ADVISORY BOARD FEEDBACK

On September 20, 2021, the core consortium of ONTOCHAIN held a virtual meeting with the members of the Advisory Board to discuss the current status of ONTOCHAIN in terms of topics procured in the first open call, the software contributions of the selected third-party projects, the overall ONTOCHAIN vision and in particular in relation to the subsequent open calls of the project. Six out of the eight members of the Advisory Board joined this meeting. This meeting was intended to provide us some feedback on the overall qualities or performance that the ONTOCHAIN platform should offer to the future applications for its end-users as a whole and the impact of these qualities to the functionality to be addressed in the topics of the subsequent open calls of ONTOCHAIN. After 40 minutes of presentations, a 20-minutes discussion was conducted on the aforementioned topics. In the open discussion, we stated as overall performance goals of the platform, the low latency, the high transaction rate, the high accuracy of trustworthiness assessment and the low transaction cost. It was pointed out by one member of the Advisory Board that ONTOCHAIN is currently missing some specific security guarantees for end-users. Moreover, one Advisory Board member pointed out that the overall greenness of the platform should be of primary concern. Then, we provided to the Advisory Board members with a short questionnaire to fill it in their time convenience, having in mind our presentation, the description of open call 2 and the recordings of our joint virtual meeting. Four out of the eight members of the Advisory Boards have participated in this survey. The results of this survey are described and discussed below.

Question 1

In the question "Do you think that the vision of ONTOCHAIN is broad enough and appropriate to respond to today's Internet challenges?", the answers obtained are depicted in Figure 17 below.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

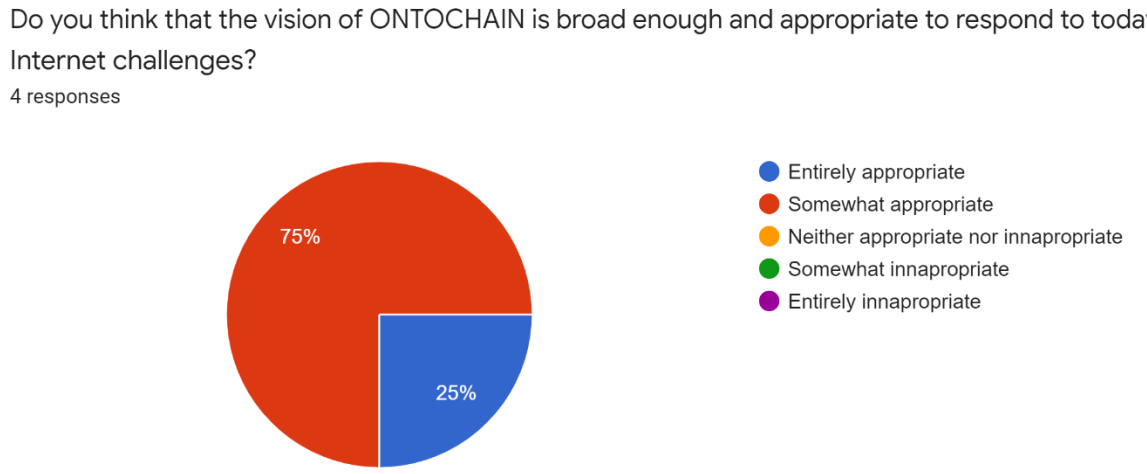

FIGURE 17: DO YOU THINK THAT THE VISION OF ONTOCHAIN IS BROAD ENOUGH AND APPROPRIATE TO RESPOND TO TODAY'S INTERNET CHALLENGES

Then, the survey was asking "*If you didn't answer 'Entirely appropriate' to the above question, then which crucial functionality is missing from ONTOCHAIN vision in your opinion?*"

Three responses were obtained to this question. 1) One considers that the security is not a fundamental element of the infrastructure design of ONTOCHAIN, which should not be the case in trustless distributed systems. 2) Another considers that ONTOCHAIN may face integration issues for the various software solution that it encapsulates. Moreover, that ONTOCHAIN relies too much on specific technological choices, so that its functionality cannot be agnostic to the specific underlying distributed ledger technology employed, which is the Ethereum. 3) A third member pointed out that ONTOCHAIN should take into account needs coming from applications addressing green and ecological transition, energy saving and critical resource management of the planet.

ONTOCHAIN key take-away messages:

1) Security and trust are fundamental design goals for ONTOCHAIN: secure data analytics without data disclosure (KnowledgeX subproject), trustworthy IPR management (Copyrightly subproject), decentralized reputation system with unlinkable feedback (REPUTABLE subproject), SSIs with verifiable credentials (OntoSSIVault subproject), SSIs bridging to real identities (HIBI subproject) are some of the functionalities that have been addressed by selected subprojects in open call 1. Therefore, while not explicitly

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

mentioned, security and trust have been the focus of the work so far in ONTOCHAIN.

2) Integration among the various different technological solutions provided by different subprojects is achieved by the fact that each different functionality is deployed as a service available on top of the Ethereum blockchain. While this facilitates integration, it makes the ONTOCHAIN platform non-DLT agnostic. The ONTOCHAIN consortium strongly believes on the scalability of L2 solutions of Ethereum and in order to minimize the time to the market, it will utilize Ethereum as its baseline blockchain technology and a L2 solution as its mainnet.
3) The interest towards applications that support the green and ecological transition will be taken into account in the third year of the project (open call 3), where different dapps that employ the ONTOCHAIN functionality will be procured.

Question 2

In the question "*Which do you think are ONTOCHAIN's main achievements?*", we received three responses: "One of the main achievements of the project is that it builds a whole ecosystem of services which interoperate for a common goal" was one answer. "The idea to provide a framework that makes the development of the next dapps easier but also, and more importantly, safer, which could also increase the trust toward this technology" was a second answer. "Providing advancement over several aspects of the blockchain, not just the cryptographic one" was the third aspect.

ONTOCHAIN key take-away messages:

The feedback obtained by the Advisory Board members in this question shows that ONTOCHAIN is in a good path to achieve its ambitious goals. The perception of the ONTOCHAIN platform facilitating trustworthy dapp development is of particular interest and we will further boost this capability of the platform in the future.

Question 3

In the question "*Which performance or other properties should ONTOCHAIN consider, apart from high transaction throughput, low latency, low cost (which may depend on the choice of the underlying ledger of ONTOCHAIN)?*", we obtained three responses as follows:

One responder considered an important property to be the temporal consistency of transactions with respect to their order of execution. Moreover, the immediate finality, as opposed to the current eventual finality offered by Ethereum or Bitcoin, was considered another important property that the ONTOCHAIN platform should possess. These properties

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

are considered of particular importance for financial applications. Moreover, it is mentioned that there is a trade-off between verifiability of information and privacy based on the data that is stored on the chain, i.e., the more on-chain data, the better verifiability, but also the less privacy. Another responder considered an important property of the system to be security. A third responder suggested to collect requirements from applications towards green and ecological transition.

ONTOCHAIN key take-away messages:

Different applications may have their own requirements and definitely eliciting the requirements of financial or green applications is something to properly look into and plan in the next open calls of the project.

Question 4

In the question "Which are the killer apps for demonstrating ONTOCHAIN competitiveness or the apps that it should definitely support?", we obtained two responses, specifically:

One responder considered that transparency in supply chain management could be a killer app. Another responder mentions that killer apps could emerge in the domain of green and ecological transition.

ONTOCHAIN key take-away messages:

Supply chain management is one of the applications that indeed were already in the plans to be offered on top of ONTOCHAIN platform. To this end, the subproject POC4Commerce, which was procured in the first open call of ONTOCHAIN, defined an ontology for annotating transactions/services/processes in the supply chain. Regarding applications on green and ecological transition, we will investigate deeper whether applications in this domain can be the suitable demonstrators for the functionality of ONTOCHAIN.

Question 5

In the question "Do you have a specific suggestion for technological choices for ONTOCHAIN that could boost its performance and/or adoption?", two responses were obtained.

One responder argued that ONTOCHAIN is (allegedly) not known outside ONTOCHAIN partners and suggested to pursue software development in cooperation with mainstream blockchain platforms through their open calls. The second responder suggested to check transition to Tendermint\Cosmos and Algorand, because of their immediate finality and

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

high transaction throughput, despite the fact that their VM is less stable than Ethereum VM (EVM).

ONTOCHAIN key take-away messages:

Participating in the open calls of different blockchain platforms would not be compliant to the funding framework of ONTOCHAIN. Although, we maintain links to Tezos foundation and GAIA-X for exchanging views, while iExec, which is a core partner of ONTOCHAIN, offers commercial blockchain services and provides deeper insights on the technological trends. It is definitely not true that ONTOCHAIN is not well known. Many applicants in the ONTOCHAIN open calls (selected on not by ONTOCHAIN) have received grants for software development from mainstream blockchain platforms. Moreover, a large number of top research groups and top SMEs in the blockchain domain (even beyond European borders) apply in our open calls to work with ONTOCHAIN.

Regarding adopting a different technological solution, other than Ethereum, EVM stability as well as the upcoming transition of Ethereum to PoS and L2 solutions, tell us that our choice to select Ethereum as our underlying blockchain technology was correct. Moreover, fintech applications, NFT markets, etc. are all deployed on top of Ethereum. In fact, any advantage offered by alternative blockchain platforms can be considered more like a promise than a really-proven technological solution in the field. In any case, migrating ONTOCHAIN platform to a different blockchain network, while requiring work, it is quite feasible, if it is deemed as necessary in the future.

Question 6

In the question "From 1 to 5, do you find the business model presented for ONTOCHAIN appropriate for the sustainability and the business impact of the project?", with 1 being "Entirely Appropriate" and 5 being "Not appropriate at all", we received the answers as depicted in Figure 18 below.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

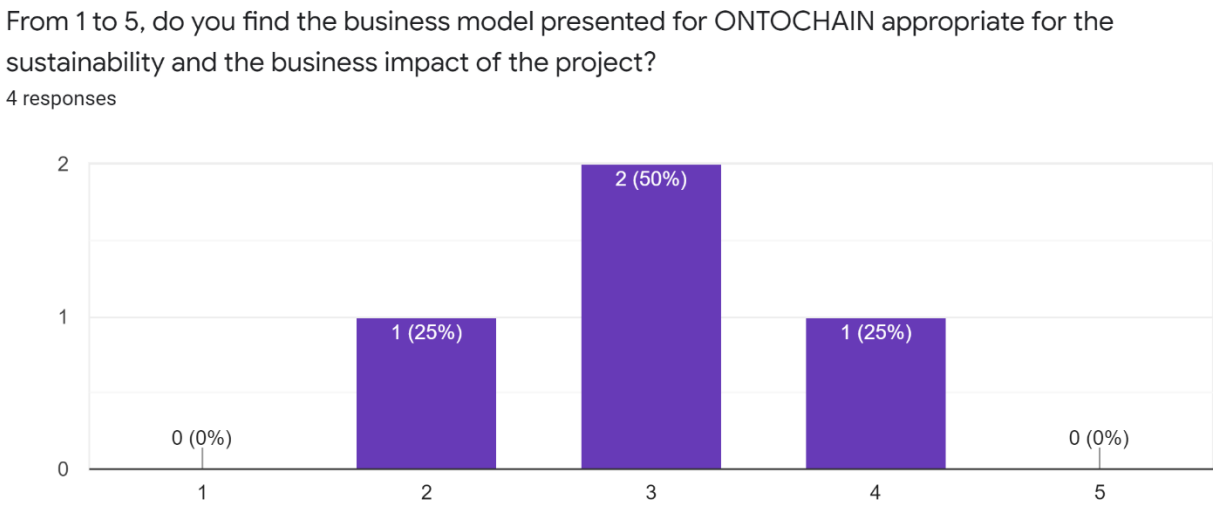

FIGURE 18: DO YOU FIND THE BUSINESS MODEL PRESENTED FOR ONTOCHAIN APPROPRIATE FOR THE SUSTAINABILITY AND THE BUSINESS IMPACT OF THE PROJECT

ONTOCHAIN key take-away messages:

The Advisory Board members clearly were not sure if the early business model presented to them is appropriate or not. This may be due to the fact that the project currently is at an early stage of development. Therefore, an elaborate business model for ONTOCHAIN, such as the one presented to them, might seem to be wishful thinking for them.

Question 7

In the question "For each ONTOCHAIN stakeholder, how do you estimate the fitness of the presented business model? (from 1 "Entirely satisfactory" to 5 "Entirely unsatisfactory")", we collected answers are summarized in Figure 19 below.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

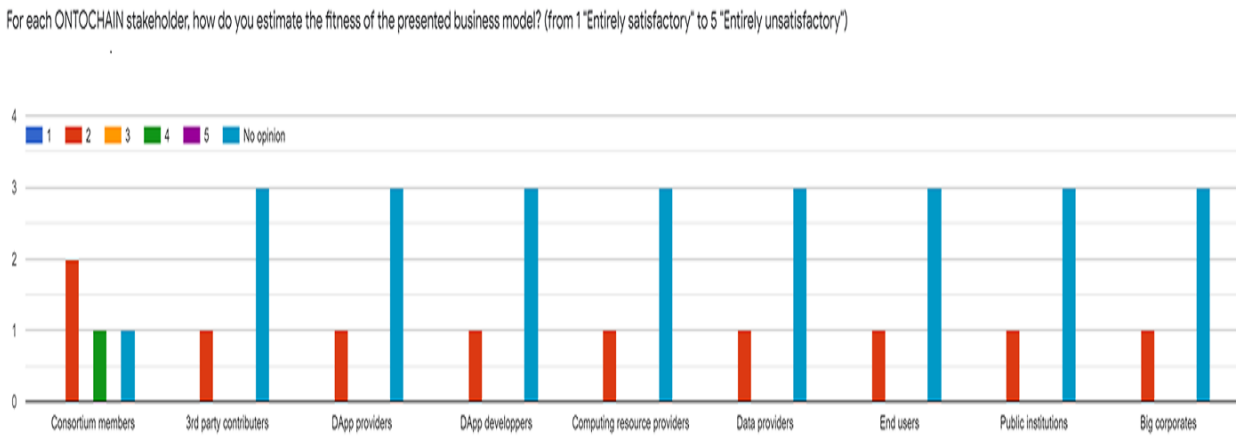

FIGURE 19: HOW DO YOU ESTIMATE THE FITNESS OF THE ONTOCHAIN BUSINESS MODEL

One responder found the business model of ONTOCHAIN satisfactory for all stakeholders, while two responders consider satisfactory for ONTOCHAIN consortium. The rest of the responders did not express any opinion.

In the follow-up question "Is there any specific remark that you would like to share about your answers to the previous question?", one responder found that the business model looks quite fair and in line with the current approaches, while another one claimed the lack of knowledge on the different stakeholders to make a right judgement for this question.

ONTOCHAIN key take-away messages:

Most of the members of the Advisory Board are not sure that the early business model of ONTOCHAIN is realistic and well-balanced for all stakeholders of the ONTOCHAIN ecosystem, while one responder is convinced on that. Again, this may be due to the fact that the project currently is at an early stage of development. Another reason could be, as stated by one responder, the lack of enough information for the interests of the different stakeholders to properly answer this question.

Question 8

Finally, in the question "Do you think that a utility token should be employed by ONTOCHAIN?", we received four responses, as graphically depicted below.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

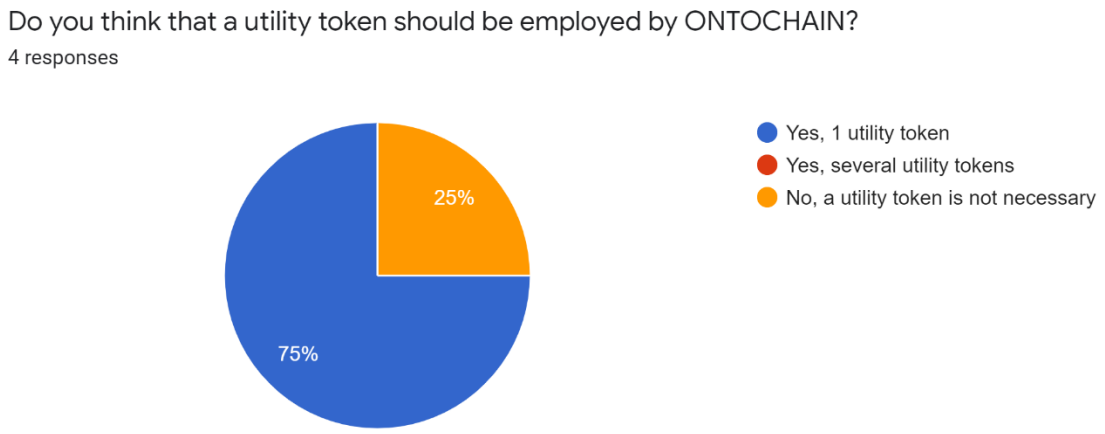

FIGURE 20: DO YOU THINK THAT A UTILITY TOKEN SHOULD BE EMPLOYED BY ONTOCHAIN

In the follow-up question "Please briefly explain your previous answer", three responses were received. One responder suggested the use of one utility token in ONTOCHAIN, because such an approach has been employed by many successful projects. Another responder did not have a strong opinion on that, but she/he thinks that it might be useful. The third responder claimed that the question did not clearly defined the context and the purpose of using utility tokens.

ONTOCHAIN key take-away messages:

Overall, most of the responders think that a utility token will be useful for ONTOCHAIN at some point. In any case, studying the employment of crypto tokens in ONTOCHAIN is planned in the next period of the project and such a token strategy (if employed) will be part of the final business model of ONTOCHAIN that will be reported in D5.4.

10. CONCLUSIONS

In this report, we reviewed the state of the art in the main functional blocks that are envisioned to jointly constitute the ONTOCHAIN framework. Note that, while the overview of the respective research domain and technologies might be quite extensive, it should only be perceived as a summary of crucial technologies that are considered within the ONTOCHAIN project and not as a thorough and exhaustive list of them. Based on the fact that blockchain semantics lie in the hurt of this framework, we separately overviewed fundamental concepts and technologies in the domains of blockchain and the semantic web. Moreover, we discussed

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

the main verticals and sample applications that could exploit the potential of ONTOCHAIN in the future. The standardization efforts and our participation in those activities were also outlined. Finally, we discussed the main technological challenges towards the realization of the ONTOCHAIN framework and the feedback received by the Advisory Board on the progress of our project and the desirable performance properties of the envisioned software. Overall, the ONTOCHAIN progress was perceived quite positively by the members of the Advisory Board and they argued that the features of the ONTOCHAIN framework should be focused around user-perceived security, privacy and trust guarantees, additionally to low latency, high throughput and low transaction cost.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

REFERENCES

Advait Deshpande, Katherine Stewart, Louise Lepetit, Salil Gunashekar (2017) Distributed Ledger Technologies/Blockchain: Challenges, Opportunities and the prospects for Standards, British Standard Institute.

Al Breiki, H., Al Qassem, L., Salah, K., Rehman, M. H. U., & Sevtinovic, D. (2019) Decentralized access control for iot data using blockchain and trusted oracles. IEEE International Conference on Industrial Internet (ICII).

Al-Breiki, H., Rehman, M. H. U., Salah, K., & Svetinovic, D.,Trustworthy blockchain oracles: review, comparison, and open research challenges. IEEE Access, 8, 85675- 85685, 2020

Alzubaidi, A., Mitra, K., Patel, P. and Solaiman, E. (2020) A Blockchain-based Approach for Assessing Compliance with SLA-guaranteed IoT Services. In IEEE International Conference on Smart Internet of Things (SmartIoT) (pp. 213-220). IEEE.

Ante, L., 2021, Smart contracts on the blockchain – A bibliometric analysis and review, Telematics and Informatics, 57.

Antonopoulos, A. (2017) Mastering Bitcoin, Publisher: O. Reilly. ISBN: 9781491954386

Ashraf, J., Hussain, O.K. and Hussain, F.K., 2014. Empirical analysis of domain ontology usage on the Web: eCommerce domain in focus. Concurrency and Computation: Practice and Experience, 26(5), pp.1157-1184.

Astraea: A decentralized blockchain oracle. Adler, J., Berryhill, R., Veneris, A., Poulos, Z., Veira, N., & Kastania, A. (2018) EEE international conference on internet of things (IThings) and IEEE green computing and communications (GreenCom) and IEEE cyber, physical and social computing (CPSCom) and IEEE smart data (SmartData) .

Aswini, A.P., 2021, "Semantic and Blockchain Technology," Advanced Concepts, Methods, and Applications in Semantic Computing, 50-71.

Azad, M.A., Bag, S. and Hao, F., 2018. PrivBox: Verifiable decentralized reputation system for online marketplaces. Future Generation Computer Systems, 89, pp.44-57.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Azad, M.A., Bag, S., Tabassum, S. and Hao, F., 2017. privy: Privacy Preserving Collaboration Across Multiple Service Providers to Combat Telecom Spams. IEEE transactions on emerging topics in computing, 8(2), pp.313-327.

Baader, F., Horrocks, I., Lutz, C. and Sattler, U., 2017. Introduction to description logic. Cambridge University Press.

Bag, S., Azad, M.A. and Hao, F., 2018. A privacy-aware decentralized and personalized reputation system. Computers & Security, 77, pp.514-530.

Belchior, R., Vasconcelos, A., Guerreiro, S. & Correia, M., 2022, A Survey on Blockchain Interoperability: Past, Present, and Future Trends, ACM Computing Surveys, 54(8).

Bethencourt, J., Sahai, A. and Waters, B., 2007, May. Ciphertext-policy attribute-based encryption. In 2007 IEEE symposium on security and privacy (SP'07) (pp. 321-334). IEEE.

Birbeck A., 2011. Introduction to RDFa. A List Apart, Zuletzt eingesehen am 06.04.2011.

BitNews (2020) "Crypto Hack: How Much Have Projects Lost In Attack," Bit News. https://en.bit.news/crypto-hack-how-much-have-projects-lost-inattack/ (accessed Apr. 23, 2021).

Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H.B., Patel, S., Ramage, D., Segal, A. and Seth, K., 2017, October. Practical secure aggregation for privacy-preserving machine learning. In proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (pp. 1175-1191).

Broder, A. and Mitzenmacher, M., 2004. Network applications of bloom filters: A survey. Internet mathematics, 1(4), pp.485-509.

Cai, C., Weng, J., Yuan, X. & Wang, C., 2021, "Enabling Reliable Keyword Search in Encrypted Decentralized Storage with Fairness," IEEE Transactions on Dependable and Secure Computing, 18(1), 131-144.

Cai, C., Weng, J., Yuan, X. & Wang, C., 2021, "Enabling Reliable Keyword Search in Encrypted Decentralized Storage with Fairness," IEEE Transactions on Dependable and Secure Computing, 18(1), 131-144.

Caldarelli, G., 2020. Understanding the blockchain oracle problem: a call for action. Information, 11(11), p.509

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Cano-Benito, J., Cimmino, A. & García-Castro, R., 2019, Towards Blockchain and Semantic Web, Lecture Notes in Business Information Processing, vol. 373 LNBIP, 220-231, Springer.

Cantone, D., Longo, C.F., Asmundo, M.N., Santamaria, D.F. and Santoro, C., 2019, June. Towards an Ontology-Based Framework for a Behavior-Oriented Integration of the IoT. In WOA (pp. 119-126).

Cantone, D., Longo, C.F., Nicolosi-Asmundo, M., Santamaria, D.F. and Santoro, C., 2020. Ontological Smart Contracts in OASIS: Ontology for Agents, Systems, and Integration of Services.

Chaum, D., Crépeau, C. and Damgard, I., 1988, January. Multiparty unconditionally secure protocols. In Proceedings of the twentieth annual ACM symposium on Theory of computing (pp. 11-19).

Chen, Y., Xie, H., Lv, K., Wei, S. and Hu, C., 2019. DEPLEST: A blockchain-based privacy-preserving distributed database toward user behaviors in social networks. Information Sciences, 501, pp.100-117.

Corda R3 Documentation (2020) Oracles. [online] Available at: https://docs.r3.com/en/platform/corda/4.8/open-source/key-concepts-oracles.html [Accessed 7 Nov. 2021].

Corrigan-Gibbs, H. and Boneh, D., 2017. Prio: Private, robust, and scalable computation of aggregate statistics. In 14th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 17) (pp. 259-282).

Cyganiak, R., Wood, D., Lanthaler M. (2014), RDF 1.1 Concepts and Abstract Syntax. Technical report, W3C.

Daskal E., Wentrup R., Shefet D. (2020). Taming the Internet Trolls with an Internet Ombudsperson: Ethical Social Media Regulation. Policy and Internet, 12(2):207-224. DOI: 10.1002/poi3.227

De Pedro, A. S., Levi, D., & Cuende, L. I. (2017) Witnet: A decentralized oracle network protocol, arXiv preprint arXiv:1711.09756.

Dean, A. and Jim, H., 2011. Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL.

Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

DIF (2021) "DIF - Decentralized Identity Foundation," identity.foundation. https://identity.foundation/ (accessed Apr. 21, 2021).

Dobre R.A., Preda R.O., Badea R.A., Stanciu M., Brumaru A. (2020). Blockchain-Based Image Copyright Protection System using JPEG Resistant Digital Signature. In: Conference Proceedings of the IEEE 26th International Symposium for Design and Technology in Electronic Packaging, SIITME 2020, pp. 206-210. DOI: 10.1109/SIITME50350.2020.9292296

Doerr, M., Ore, C.E. and Stead, S., 2007, November. The CIDOC conceptual reference model: a new standard for knowledge sharing. In Tutorials, posters, panels and industrial contributions at the 26th international conference on Conceptual modeling-Volume 83 (pp. 51-56).

Dou, Y., Chan, H.C. and Au, M.H., 2018. A distributed trust evaluation protocol with privacy protection for intercloud. IEEE Transactions on Parallel and Distributed Systems, 30(6), pp.1208-1221.

DuCharme, B., 2011 Learning SPARQL. O'Reilly Media, Inc.

Dürst, M. and Suignard, M., 2005. Internationalized resource identifiers (IRIs) (p. 8). RFC 3987, January.

Dwork, C., McSherry, F., Nissim, K. and Smith, A., 2006, March. Calibrating noise to sensitivity in private data analysis. In Theory of cryptography conference (pp. 265-284). Springer, Berlin, Heidelberg.

Ellis, S., Juels, A. and Nazarov, S., 2017. Chainlink: A decentralized oracle network. Retrieved March, 11, p.2018.

English, M., Auer, S. & Domingue, J., 2016, Block Chain Technologies & The Semantic Web: A Framework for Symbiotic Development. Computer Science Conference for University of Bonn Students.

Erkin, Z., Veugen, T., Toft, T. and Lagendijk, R.L., 2012. Generating private recommendations efficiently using homomorphic encryption and data packing. IEEE transactions on information forensics and security, 7(3), pp.1053-1066.

Erlingsson, Ú., Pihur, V. and Korolova, A., 2014, November. Rappor: Randomized aggregatable privacy-preserving ordinal response. In

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Proceedings of the 2014 ACM SIGSAC conference on computer and communications security (pp. 1054-1067).

Esteves, B., & Rodríguez-Doncel, V. (2021). Analysis of Ontologies and Policy Languages to Represent Information Flows in GDPR, submitted to Semantic Web Journal (2703-3917). http://www.semantic-web- journal.net/system/files/swj2703.pdf.

Eternity blockchain. Hess, Z., Malahov, Y., & Pettersson, J. (2017) Available online: https://aeternity. com/aeternity-blockchainwhitepaper.pdf

ETSI TS 103 458 V1.1.1 (2018) "CYBER; Application of Attribute Based Encryption (ABE) for PII and personal data protection on IoT devices, WLAN, cloud and mobile services - High level requirements". (https://www.etsi.org/deliver/etsi_ts/103400_103499/103458/01.01.01_60/ts_103458v010101p.pdf)

ETSI TS 103 532 V1.1.1, (2018) "CYBER; Attribute Based Encryption for Attribute Based Access Control. (https://www.etsi.org/deliver/etsi_ts/103500_103599/103532/01.01.01_60/ts_103532v010101p. pdf)

Ferdous, Md. Sadek, Mohammad Jabed Morshed Chowdhury, Mohammad Ashraful Hoque and Alan W. Colman (2020) "Blockchain Consensus Algorithms: A Survey." arXiv: Distributed, Parallel, and Cluster Computing.

Fortino, G., Messina, F., Rosaci, D. and Sarné, G.M., 2019. Using blockchain in a reputation-based model for grouping agents in the Internet of Things. IEEE Transactions on Engineering Management, 67(4), pp.1231-1243.

Fredrikson, M., Lantz, E., Jha, S., Lin, S., Page, D. and Ristenpart, T., 2014. Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing. In 23rd {USENIX} Security Symposium ({USENIX} Security 14) (pp. 17-32).

Gan, X., Li, Y., Huang, Y., Fu, L. and Wang, X., 2019. When crowdsourcing meets social IoT: An efficient privacy-preserving incentive mechanism. IEEE Internet of Things Journal, 6(6), pp.9707-9721.

Gangemi, A., Guarino, N., Masolo, C., Oltramari, A. and Schneider, L., 2002, October. Sweetening ontologies with DOLCE. In International Conference on Knowledge Engineering and Knowledge Management (pp. 166-181). Springer, Berlin, Heidelberg.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

García R., Gil R. (2019). Social media copyright management using semantic web and blockchain. In: ACM International Conference Proceeding Series. DOI: 10.1145/3366030.3366128

Glimm, B., Horrocks, I., Motik, B., Stoilos, G. and Wang, Z., 2014. HermiT: an OWL 2 reasoner. Journal of Automated Reasoning, 53(3), pp.245-269

gnosis.io. (n.d.). Gnosis. [online] Available at: https://gnosis.io/ [Accessed 7 Nov. 2021].

Goldmann, N., 2021, E-commerce. Journal of Internet Banking and Commerce

Gómez-Pérez, A., 2004. Ontology evaluation. In Handbook on ontologies (pp. 251-273). Springer, Berlin, Heidelberg.

Gray, Marley. (n.d.). Introducing Project Bletchley and elements of blockchain born in the Microsoft Cloud. [online] Available at: https://azure.microsoft.com/en-us/blog/bletchley-blockchain/ [Accessed 7 Nov. 2021].

Guha, R., McCool, R. and Miller, E., 2003, May. Semantic search. In Proceedings of the 12th international conference on World Wide Web (pp. 700-709).

Gupta, M., Judge, P. and Ammar, M., 2003, June. A reputation system for peer-to-peer networks. In Proceedings of the 13th international workshop on Network and operating systems support for digital audio and video (pp. 144-152).

Halevi, S., Lindell, Y. and Pinkas, B., 2011, August. Secure computation on the web: Computing without simultaneous interaction. In Annual Cryptology Conference (pp. 132-150). Springer, Berlin, Heidelberg.

Hang, L. and Kim, D.H. (2019) Sla-based sharing economy service with smart contract for resource integrity in the internet of things. Applied Sciences, 9(17), p.3602.

He, K., Fan, H., Wu, Y., Xie, S. and Girshick, R., 2020. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 9729-9738).

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Hector, U.-R. & Boris, C.-L., 2020, "BLONDiE: Blockchain Ontology with Dynamic Extensibility."

Hendrikx, F., Bubendorfer, K. and Chard, R., 2015. Reputation systems: A survey and taxonomy. Journal of Parallel and Distributed Computing, 75, pp.184-197.

Hepp, M., 2008, September. Goodrelations: An ontology for describing products and services offers on the web. In International conference on knowledge engineering and knowledge management (pp. 329-346). Springer, Berlin, Heidelberg.

Hewa, T., Ylianttila, M. & Liyanage, M., 2021, Survey on blockchain based smart contracts: Applications, opportunities and challenges, Journal of Network and Computer Applications, 177.

Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P.F. and Rudolph, S., 2009. OWL 2 web ontology language primer. W3C recommendation, 27(1), p.123.

Hofweber, T., (2018) Logic and Ontology. Edward N. Zalta (ed.), The Stanford Encyclopaedia of Philosophy.

Identity Management Institute (2021) "Self-Sovereign Identity - Identity Management Institute." https://www.identitymanagementinstitute.org/self-sovereignidentity/ (accessed Apr. 23, 2021).

Ion, M., Zhang, J. and Schooler, E.M., 2013, August. Toward content-centric privacy in ICN: Attribute-based encryption and routing. In Proceedings of the 3rd ACM SIGCOMM workshop on Information-centric networking (pp. 39-40).

IRFT RFC7954 (2016) "Information-Centric Networking: Evaluation and Security Considerations", https://datatracker.ietf.org/doc/html/rfc7954

IRTF RFC7476 (2015) "Information-Centric Networking: Baseline Scenarios", https://datatracker.ietf.org/doc/html/rfc7476

Jin, X. and Zhang, Y., 2018. Privacy-preserving crowdsourced spectrum sensing. IEEE/ACM Transactions on Networking, 26(3), pp.1236-1249.

Jøsang, A., Ismail, R. and Boyd, C., 2007. A survey of trust and reputation systems for online service provision. Decision support systems, 43(2), pp.618-644.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Jovanovic, J. and Bagheri, E., 2016. Electronic commerce meets the semantic web. It Professional, 18(4), pp.56-65.

Kamvar, S.D., Schlosser, M.T. and Garcia-Molina, H., 2003, May. The eigentrust algorithm for reputation management in p2p networks. In Proceedings of the 12th international conference on World Wide Web (pp. 640-651).

Kim, J.W., Kim, D.H. and Jang, B., 2018. Application of local differential privacy to collection of indoor positioning data. Ieee Access, 6, pp.4276-4286.

Kinateder, M. and Pearson, S., 2003, September. A privacy-enhanced peer-to-peer reputation system. In International Conference on Electronic Commerce and Web Technologies (pp. 206-215). Springer, Berlin, Heidelberg.

Klyne, G., 2004. Resource description framework (RDF): Concepts and abstract syntax. http://www. w3. org/TR/2004/REC-rdf-concepts-20040210/.

Kochovski, P., Stankovski, V., Gec, S., Faticanti, F., Savi, M., Siracusa, D. and Kum, S., 2020. Smart contracts for service-level agreements in edge-to-cloud computing. Journal of Grid Computing, 18(4), pp.673-690.

Konashevych O. (2020). Constraints and benefits of the blockchain use for real estate and property rights. Journal of Property, Planning and Environmental Law, 12(2):109-127. DOI: 10.1108/JPPEL-12-2019-0061

Krause, E., 2018. A Fifth of All Bitcoin Is Missing. These Crypto Hunters Can Help. Wall Street Journal. Archived from the original on, 9.

Kripa M., Nidhin Mahesh A., Ramaguru R., Amritha P.P. (2021). Blockchain Framework for Social Media DRM Based on Secret Sharing. In: Smart Innovation, Systems and Technologies, 195, pp. 451-458. DOI: 10.1007/978-981-15-7078-0_43

Lemieux, V.L., 2017, November. Blockchain and distributed ledgers as trusted recordkeeping systems. In Future Technologies Conference (FTC) (Vol. 2017).

Lesavre, L., Varin, P. and Yaga, D., 2020. Blockchain Networks: Token Design and Management Overview (No. NIST Internal or Interagency Report (NISTIR) 8301 (Draft)). National Institute of Standards and Technology.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Li, K., Tian, L., Li, W., Luo, G. and Cai, Z., 2019. Incorporating social interaction into three-party game towards privacy protection in IoT. Computer Networks, 150, pp.90-101.

Liang, X., Shetty, S., Tosh, D., Kamhoua, C., Kwiat, K. and Njilla, L., 2017, May. Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability. In 2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID) (pp. 468-477). IEEE.

Liu, D., Alahmadi, A., Ni, J., Lin, X. and Shen, X., 2019. Anonymous reputation system for IIoT-enabled retail marketing atop PoS blockchain. IEEE Transactions on Industrial Informatics, 15(6), pp.3527-3537.

Liu, X., and Feng, J. (2021) Trusted Blockchain Oracle Scheme Based on Aggregate Signature. Journal of Computer and Communications.

López-Pimentel, J.C., Rojas, O. and Monroy, R., 2020, November. Blockchain and off-chain: A Solution for Audit Issues in Supply Chain Systems. In 2020 IEEE International Conference on Blockchain (Blockchain) (pp. 126-133). IEEE.

Lu, Z., Liu, W., Wang, Q., Qu, G. and Liu, Z., 2018. A privacy-preserving trust model based on blockchain for VANETs. IEEE Access, 6, pp.45655-45664.

Ma, L., Kaneko, K., Sharma, S., & Sakurai, K. (2019) Reliable decentralized oracle with mechanisms for verification and disputation. Seventh International Symposium on Computing and Networking Workshops (CANDARW).

Makni, B., Abdelaziz, I. and Hendler, J., 2020. Explainable Deep RDFS Reasoner. arXiv preprint arXiv:2002.03514.

Manola, F., Miller, E., (2004). RDF Primer. W3C Recommendation. World Wide Web Consortium.

Masinter, L., Berners-Lee, T. and Fielding, R.T., 2005. Uniform resource identifier (URI): Generic syntax. Network Working Group: Fremont, CA, USA.

Melis, L., Danezis, G. and De Cristofaro, E., 2016. Efficient private statistics with succinct sketches. NDSS.

Miao, C., Jiang, W., Su, L., Li, Y., Guo, S., Qin, Z., Xiao, H., Gao, J. and Ren, K., 2019. Privacy-preserving truth discovery in

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

crowd sensing systems. ACM Transactions on Sensor Networks (TOSN), 15(1), pp.1-32.

Microsoft (2021a) "Decentralized Identity, Blockchain, and Privacy | Microsoft Security," Microsoft Security. https://www.microsoft.com/enus/security/business/identity-access-management/decentralized-identityblockchain (accessed Apr. 21, 2021).

Microsoft (2021b) "Decentralized Identity: Own and control your identity." Accessed: Apr. 23, 2021. [Online]. Available: https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE2DjfY.

Mikroyannidis, A., Third, A. & Domingue, J., 2020, "A case study on the decentralisation of lifelong learning using blockchain technology," Journal of Interactive Media in Education, 2020(1), 1–10.

Mohsen, W., Aref, M. & Elbahnasy, K., 2020, Blockchain as a Platform for Collaborative Ontology Evolution, PervasiveHealth: Pervasive Computing Technologies for Healthcare, 183–190, ICST.

Moin, S., Karim, A., Safdar, Z., Safdar, K., Ahmed, E. and Imran, M., 2019. Securing IoTs in distributed blockchain: Analysis, requirements and open issues. Future Generation Computer Systems, 100, pp.325-343.

Nakamoto, S., no date, Bitcoin: A Peer-to-Peer Electronic Cash System.

Nakashima, H. and Aoyama, M., 2017, June. An automation method of sla contract of web apis and its platform based on blockchain concept. In 2017 IEEE International Conference on Cognitive Computing (ICCC) (pp. 32-39). IEEE.

Nasikas, C.S., 2018. Accountable and privacy preserving data processing via distributed ledgers.

Nijsse, Jeff, and Litchfield, Alan (2020) "A Taxonomy of Blockchain Consensus Methods" Cryptography 4, no. 4: 32. https://doi.org/10.3390/cryptography4040032

No date, "A more pragmatic Web 3.0: Linked Blockchain Data."

Oberle D., Guarino N., Staab S., (2009). What is an ontology? Handbook on Ontologies, 2nd edition. Springer.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Obrst, L., Ceusters, W., Mani, I., Ray, S. and Smith, B., 2007. The evaluation of ontologies. In Semantic web (pp. 139-158). Springer, Boston, MA.

OraclesLink: An architecture for secure oracle usage. Berger, B., Huber, S., & Pfeifhofer, S. (2020) Second International Conference on Blockchain Computing and Applications (BCCA).

Panarello, A., Tapas, N., Merlino, G., Longo, F. & Puliafito, A. (2018) Blockchain and iot integration: A systematic survey, Sensors (Switzerland), 18(8).

Pandit, H.J. (2020). Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance (Doctoral dissertation, Trinity College Dublin).

Patel-Schneider, P.F., 2014, October. Analyzing schema. org. In International Semantic Web Conference (pp. 261-276). Springer, Cham.

Pavlov, E., Rosenschein, J.S. and Topol, Z., 2004, March. Supporting privacy in decentralized additive reputation systems. In International Conference on Trust Management (pp. 108-119). Springer, Berlin, Heidelberg.

Pérez, S., Hernández-Ramos, J.L., Matheu-García, S.N., Rotondi, D., Skarmeta, A.F., Straniero, L. and Pedone, D., 2018. A lightweight and flexible encryption scheme to protect sensitive data in smart building scenarios. IEEE Access, 6, pp.11738-11750.

Peterson, J., Krug, J., Zoltu, M., Williams, A. K., & Alexander, S. Augur (2015) A decentralized oracle and prediction market platform. arXiv preprint arXiv:1501.01042.

Petkus, M. 2019 "Why and How zk-SNARK Works: Definitive Explanation"

Pham, C., Adamopoulos, A. and Tait, E., 2019. Towards a Triple Bottom Line Perspective of Blockchains in Supply Chain.

Polat, H. and Du, W., 2003, November. Privacy-preserving collaborative filtering using randomized perturbation techniques. In Third IEEE International Conference on Data Mining (pp. 625-628). IEEE.

Primault, V., Lampos, V., Cox, I. and De Cristofaro, E., 2019, May. Privacy-preserving crowd-sourcing of web searches with

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

private data donor. In The World Wide Web Conference (pp. 1487-1497).

Protocol Labs. The Interplanetary File System (IPFS). https://ipfs.io/.

Ramachandran, A. and Kantarcioglu, D., 2017. Using blockchain and smart contracts for secure data provenance management. arXiv preprint arXiv:1709.10000.

Ramani, S.K., Tourani, R., Torres, G., Misra, S. and Afanasyev, A., 2019, September. Ndn-abs: Attribute-based signature scheme for named data networking. In Proceedings of the 6th ACM Conference on Information-Centric Networking (pp. 123-133).

Ranchal, R. and Choudhury, O., 2020, October. SLAM: A Framework for SLA Management in Multicloud ecosystem using Blockchain. In 2020 IEEE Cloud Summit (pp. 33-38). IEEE.

Rashidi, B., Fung, C., Nguyen, A., Vu, T. and Bertino, E., 2017. Android user privacy preserving through crowdsourcing. IEEE Transactions on Information Forensics and Security, 13(3), pp.773-787.

Ruta, M., Scioscia, F., Ieva, S., Capurso, G. & Sciascio, E. di (2017a) "Regular research paper: Blockchain, Service-Oriented Architecture, Semantic Web of Things, Semantic Web," Journal of Internet of Things (OJIOT), 3(1).

Sakemi, Y., Kobayashi, T., Saito, T. and Wahby, R.S., 2020. Pairing-friendly curves. Internet Engineering Task Force, Internet-Draft draft-irtf-cfrg-pairing-friendly-curves-05.

Sandhu, R.S. and Samarati, P., 1994. Access control: principle and practice. IEEE communications magazine, 32(9), pp.40-48.

Sandhu, R.S., Coyne, E. J., Feinstein, H. L., Youman C. E., 1996 "Role-based access control models", Computer, vol. 29, no. 2, pp. 38-47.

Scheid, E.J. and Stiller, B. (2018) Automatic SLA Compensation based on Smart Contracts. Technical Report No. IFI-2018.02, April.

Shwetha, A.N. and Prabodh, C.P., 2021. Auction System in Food Supply Chain Management Using Blockchain. In Proceedings of International Conference on Advances in Computer Engineering and Communication Systems (pp. 31-40). Springer, Singapore.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Stonebraker, M. and Hellerstein, J.M., 2001, May. Content integration for e-business. In Proceedings of the 2001 ACM SIGMOD international conference on Management of data (pp. 552-560).

Tang, B., Kang, H., Fan, J., Li, Q. and Sandhu, R., 2019, May. Iot passport: A blockchain-based trust framework for collaborative internet-of-things. In Proceedings of the 24th ACM symposium on access control models and technologies (pp. 83-92).

Tartir, S., Arpinar, I.B., Moore, M., Sheth, A.P. and Aleman-Meza, B., 2005. OntoQA: Metric-based ontology quality analysis.

Tello, A. and Gómez-Pérez, A., 2004. Ontometric: A method to choose the appropriate ontology. Journal of Database Management (JDM), 15(2), pp.1-18.

Tran, N.K., Ali Babar, M. & Boan, J., 2021, Integrating blockchain and Internet of Things systems: A systematic review on objectives and designs, Journal of Network and Computer Applications, 173.

Tykn (2021) "Self-Sovereign Identity: The Ultimate Beginners Guide!," Tykn.tech. https://tykn.tech/self-sovereign-identity/ (accessed Apr. 23, 2021).

Ugarte, H., 2017. A more pragmatic Web 3.0: Linked blockchain data. Bonn, Germany.

Uriarte, R.B., De Nicola, R. and Kritikos, K., 2018, December. Towards distributed sla management with smart contracts and blockchain. In 2018 IEEE International Conference on Cloud Computing Technology and Science (CloudCom) (pp. 266-271). IEEE.

Valiente, M.-C., Rozas, D. & Hassan, S., no date, Integration of ontologies with decentralized autonomous organizations development: A systematic review.

Videnov, S., 2019. Decentralised data provenance based on the blockchain (Doctoral dissertation, Wien).

W3C (2019) "Verifiable Credentials Data Model 1.0," https://www.w3.org/TR/vc-data-model/ (accessed Apr. 21, 2021).

W3C (2021) "Decentralized Identifiers (DIDs) v1.0," w3.org. https://www.w3.org/TR/did-core/ (accessed Apr. 21, 2021).

Walsh, K. and Sirer, E.G., 2006, May. Experience with an Object Reputation System for Peer-to-Peer Filesharing. In NSDI (Vol. 6, pp. 1-1).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

Wang, Y. and Singh, M.P., 2010. Evidence-based trust: A mathematical model geared for multiagent systems. ACM Transactions on Autonomous and Adaptive Systems (TAAS), 5(4), pp.1-28.

Wang, Z., Pang, X., Chen, Y., Shao, H., Wang, Q., Wu, L., Chen, H. and Qi, H., 2018. Privacy-preserving crowd-sourced statistical data publishing with an untrusted server. IEEE Transactions on Mobile Computing, 18(6), pp.1356-1367.

Wonjiga, A.T., Peisert, S., Rilling, L. and Morin, C., 2019, December. Blockchain as a trusted component in cloud SLA verification. In Proceedings of the 12th IEEE/ACM International Conference on Utility and Cloud Computing Companion (pp. 93-100).

Woo, S., Song, J., & Park, S. (2020) A distributed oracle using intel SGX for blockchain-based IoT applications. s.l.: Sensors, Sensors, Vol. 9.

World Economic Forum (WEF) (2020)"Redesigning Trust: Blockchain Deployment Toolkit"

World Economic Forum (WEF) (2020) "Global Standards Mapping Initiative: An overview of blockchain technical standards", whitepaper, https://www.weforum.org/whitepapers/global-standards-mapping-initiative-an-overview-of-blockchain-technical-standards

World Wide Web Consortium (2004). SWRL: A Semantic Web Rule Language Combining OWL and RuleML.

World Wide Web Consortium (2008) RDFa Primer: Bridging the human and data webs.

World Wide Web Consortium (2013) SPARQL 1.1 Query Language.

World Wide Web Consortium (2014) RDF 1.1: On Semantics of RDF Datasets.

World Wide Web Consortium. Linked data platform 1.0, 2015.

World Wide Web Consortium. RDF Schema, Frebruary 2014.

Wu, H., Wang, L. and Xue, G., 2019. Privacy-aware task allocation and data aggregation in fog-assisted spatial crowdsourcing. IEEE Transactions on Network Science and Engineering, 7(1), pp.589-602.

Yamashita, K., Nomura, Y., Zhou, E., Pi, B., & Jun, S., Potential risks of hyperledger fabric smart contracts. In 2019 IEEE

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338

International Workshop on Blockchain Oriented Software Engineering (IWBOSE) (pp. 1-10). IEEE., 2019

Yang, K., Zhang, K., Ren, J. and Shen, X., 2015. Security and privacy in mobile crowdsourcing networks: challenges and opportunities. IEEE communications magazine, 53(8), pp.75-81.

Yang, R., Au, M.H., Xu, Q. and Yu, Z., 2019. Decentralized blacklistable anonymous credentials with reputation. Computers & Security, 85, pp.353-371.

You, Y., Chen, T., Sui, Y., Chen, T., Wang, Z. and Shen, Y., 2020. Graph contrastive learning with augmentations. Advances in Neural Information Processing Systems, 33, pp.5812-5823.

Yuan, D., Li, Q., Li, G., Wang, Q. and Ren, K., 2019. PriRadar: A privacy-preserving framework for spatial crowdsourcing. IEEE transactions on information forensics and security, 15, pp.299-314.

Zhang, F., Cecchetti, E., Croman, K., Juels, A., & Shi, E., Town crier: An authenticated data feed for smart contracts. In Proceedings of the 2016 aCM sIGSAC conference on computer and communications security (pp. 270-282). , 2016

Zhang, Z., He, S., Chen, J. and Zhang, J., 2018. REAP: An efficient incentive mechanism for reconciling aggregation accuracy and individual privacy in crowdsensing. IEEE Transactions on Information Forensics and Security, 13(12), pp.2995-3007.

Zheng, Y., Duan, H., Yuan, X. and Wang, C., 2017. Privacy-aware and efficient mobile crowdsensing with truth discovery. IEEE Transactions on Dependable and Secure Computing, 17(1), pp.121-133.

Zhou, H., de Laat, C. and Zhao, Z., 2018, December. Trustworthy cloud service level agreement enforcement with blockchain based smart contract. In 2018 IEEE International Conference on Cloud Computing Technology and Science (CloudCom) (pp. 255-260). IEEE.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957338