NGI ONTO CHAIN

Blockchain for the Next Generation Internet

D3.3 TECHNICAL INVENTORY, REFERENCE ARCHITECTURE AND COMPONENTS SPECIFICATION

2/11/2021

Grant Agreement No.: 957338
Call: H2020-ICT-2020-1

Topic: ICT-54-2020
Type of action: RIA

D3.3 TECHNICAL INVENTORY, REFERENCE ARCHITECTURE AND COMPONENTS SPECIFICATION

WORK PACKAGE	WP3
TASK	T3.2
DUE DATE	31/8/2021
SUBMISSION DATE	3/11/2021
DELIVERABLE LEAD	IEXEC
VERSION	1.0
AUTHORS	Anthony Simonet-Boulogne (IEXEC) Alberto Ciaramella (IS)
REVIEWERS	Petar Kochovski (UL) Thanasis Papaioannou (AUEB)
ABSTRACT	This deliverable provides a review of existing open-source components and defines the reference architecture to be considered by third parties during the implementation of ONTOCHAIN Open Call #1.
KEYWORDS	Decentralisation, blockchain, trustworthy content, data traceability, trustworthy knowledge exchange, privacy protection, web semantic, service interoperability

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 contributor(s)
0.1	2/8/2021	Table of content & initial draft	Anthony Simonet-Boulogne
0.2	4/8/2021	Executive summary	Anthony Simonet-Boulogne
0.3	10/8/2021	"Challenges & Objectives" draft	Anthony Simonet-Boulogne
0.4	18/8/2021	"Framework & Components Specification" draft	Anthony Simonet-Boulogne
0.5	1/9/2021	"Technical Inventory" draft	Anthony Simonet-Boulogne
0.6	28/9/2021	"Technical Inventory" update	Alberto Ciaramella
0.7	27/10/2021	Corrections based on reviewers' comments	Anthony Simonet-Boulogne
0.8	29/10/2021	Styling	Anthony Simonet-Boulogne
1.0	2/11/2021	Sections 4.2 & 5	Alberto Ciaramella

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

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 European Commission. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein. The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. Moreover, it is clearly stated that the ONTOCHAIN Consortium reserves the right to update, amend or modify any part, section or detail of the document at any point in time without prior information.

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

COPYRIGHT NOTICE

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

The ONTOCHAIN Consortium is the following:

Participant number	Participant organisation 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 document is the deliverable "D3.3 Technical Inventory, Reference Architecture and Components Specification" of the European project "ONTOCHAIN– Trusted, traceable and transparent ontological knowledge on blockchain" (hereinafter also referred to as "ONTOCHAIN ", project reference: 957338.

The technical inventory, the reference architecture and the components specification describe the high-level direction that will be taken during the first year of the project. This document is also intended to provide guidance to the first parties when applying to Open Call #1 (OC1), and further during the implementation phase of OC1.

The focus of this document is first to provide a review of existing open-source and free-software components relevant to the design of the ONTOCHAIN ecosystem, then to define a reference architecture for the ONTOCHAIN ecosystem and finally to specify which components must be produced by the third parties selected through the open calls for participation. The intended audience of D3.3 is thus quite large as it consists of the ONTOCHAIN consortium, of the Project Officer and of the broader European community of innovators that will be targeted by OC1.

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

LIST OF FIGURES

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

ABBREVIATIONS

APIs	Application Programming Interfaces
DLT	Distributed Ledger Technology
NGI	Next Generation Internet
OC	Open Call for participation
OWL	Web Ontology Language
RDF	Resource Description Framework
SDK	Software Development Kit
SPARQL	SPARQL Protocol and RDF Query Language

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

1 INTRODUCTION

The ONTOCHAIN project aims to marry Semantic Web with Blockchains in order to provide a framework and ecosystem for the trusted exchange of information and services in the Next Generation Internet (NGI). The NGI initiative promotes values that are critical for an Internet for humans such as openness, inclusivity, transparency, privacy, cooperation, and protection of data. To generate this ecosystem, ONTOCHAIN offers innovators up to 4.2 Million funding through 3 Open Calls for participation, as well as mentoring from top international experts in semantic web, linked data, ontology engineering, blockchain, knowledge management, distributed and decentralized computing, business models and blockchain economics.

The purpose of this document is to provide applicants to Open Call #1 (OC1) and the selected third parties critical information regarding design requirements and recommended building blocks available publicly as open-source software or as free software for building the envisioned ONTOCHAIN ecosystem. Moreover, this deliverable describes the high-level architecture of the future ONTOCHAIN ecosystem and a description of its individual components. While the selected third parties will be working remotely on individual components, this document should be considered as a guide for ensuring the proper integration of the components in the next phases of the project.

The remainder of this deliverable is organized as follows:

- In chapter 2, we overview ONTOCHAIN's challenges and strategic objectives to be considered by all the parties involved in the conception and in the implementation of the project;
- In chapter 3, we provide a technical inventory of significant open-source and free-software building blocks that have already been adopted by the industry and can serve as a baseline for the implementation of ONTOCHAIN;
- In chapter 4 we describe the architecture of the envisioned ONTOCHAIN framework and software ecosystem, and provide detailed information about the role of its individual components.
- In chapter 5 we provide concluding remarks to this deliverable.

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

2 ONTOCHAIN CHALLENGES & OBJECTIVES

> "The overall mission of the Next Generation Internet initiative is to re-imagine and re-engineer the Internet for the third millennium and beyond. We envision the information age will be an era that brings out the best in all of us. We want to enable human potential, mobility and creativity at the largest possible scale while dealing responsibly with our natural resources. In order to preserve and expand the European way of life, we shape a value-centric, human and inclusive Internet for all." [1]

2.1 ONTOCHAIN CHALLENGES

The aforementioned significant ambitions need a solid foundation on which people –representing different cross-sector domains– can build on; and this is getting more and more visible nowadays that the technological innovation –in some cases– lacks transparency and trustworthiness.

In order to enter into the discussion on how to technically achieve and maintain transparency and trustworthiness in the Next Generation Internet technology, let us be clear that in specific situations there may be philosophical, moral, or even proven mathematical-logical reasons for people not to be able to achieve trustworthiness. On the question of the existence of a specific truth, for example, whether a person has broken the law or not, and should be sentenced to prison for that wrongdoing, let us remind ourselves of the two incompleteness theorems, published by Kurt Gödel[2] in 1931. These are two theorems of mathematical logic that demonstrate the inherent limitations of every formal axiomatic system capable of modelling (even) basic arithmetic. The theorems are widely, but not universally, interpreted as showing that Hilbert's program to find a complete and consistent set of axioms for all mathematics is impossible. The first incompleteness theorem states that no consistent system of axioms whose theorems can be listed by an effective procedure (i.e., an algorithm, for example, such, that would execute as a Smart Contract on the blockchain) is capable of proving all truths about the arithmetic of natural numbers. For any such consistent formal system, there will always be statements about natural numbers that are true, but that are unprovable within the system. The second incompleteness theorem, an extension of the first, shows that the system cannot demonstrate its own consistency. Consequently, consistency, and therefore trustworthiness may not be computable for many complex problems of human endeavour, such as the judgement to sentence

[1] Next Generation Internet 2025, A study prepared for the European Commission DG Communications Networks, Content & Technology

[2] https://en.wikipedia.org/wiki/G%C3%B6del%27s_incompleteness_theorems

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

a person to a prison for breaking the law. That, to our advantage or disadvantage, is something that only humans can decide!

Having written that, greater transparency and trustworthiness can still be achieved in the Next Generation Internet. This proposal argues the vision and approach to ONTOCHAIN, a set of new technologies to realize formal logic (e.g. first order logic) and the ability to execute formal proofs directly on blockchain. The consortium believes that ONTOCHAIN can become a key technology block of the Next Generation Internet.

However, before elaborating the technology, let us for a moment contemplate on the human problems. It has been said that even if the final formulae of the Universe is found - all of human problems will not be solved! A recent global movement and an initiative Contract for the Web[3] was introduced by the World Wide Web Foundation[4] and led by Tim Berners-Lee. This movement bootstrapped a global action plan to save the web from political manipulation, privacy violations and fake news. It is targeted towards governments, companies and individuals to make commitments in protecting the web from abuse. The contract has been worked on by 80 organisations and outlines nine principles to safeguard the web 3 each for governments, companies and individuals. The document has the backing of more than 150 organisations, from Google, Twitter, Microsoft, Facebook, Electronic Frontier Foundation etc. Those who back the contract must show they are implementing the principles and working on solutions to the tougher problems, or face being removed from the list of endorsers. The contract's principles require government to do all they can to ensure that everyone who wants to can connect to the web and have their privacy respected. People should have access to whatever personal data is held on them and have the right to object or withdraw from having that data processed. To build trust online, companies are compelled to simplify privacy settings by providing control panels where people can access their data and manage their privacy options in one place. Another principle requires companies to assess the risk of their technology spreading misinformation or harming people's behaviour or personal wellbeing. 3 more principles call on individuals to create rich and relevant content to make the web a valuable place, build strong online communities where everyone feels safe and welcome, and finally, to fight for the web, so it remains open to everyone, everywhere.

2.1.1 Current Internet Threats

Currently, the society organization, the governance and the policies structure a framework to facilitate free speech and private enterprise; nevertheless it cannot from its current standpoint- assure that any bias or systematic abuse of global trust is avoided. Moreover, the success of the Internet lies in permission-free innovation, openness, in-

[3] https://contractfortheweb.org/
[4] https://webfoundation.org/

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

teroperability and the non-limitation of choices. At the same time, though, there are specific indications that the trade-off between openness and trustworthiness is questioned. More specifically in the real-life scenarios of persons' interaction with the Internet, the following threats summarized in the table have been identified:

- ***Centralization of power:*** Innovative ideas and uniqueness of their services offering made many popular websites such as Google, Facebook, YouTube and Amazon emerge into robust centralized platforms. Even though the Internet started as a truly decentralized network, balance of power has been broken by the dominating services that now support the Internet. The networks of today are completely centralized, with the power of information and knowledge being in the hands of only a few actors. The amount of power has made these few companies the gatekeepers of knowledge and information, who the public now needs to trust to use that power in a responsible and fair manner. Keeping the knowledge and ontologies for themselves while serving them to billions of users, the gatekeepers can easily dictate what is true and what is false.

- ***Unknown provenance of information:*** We all make daily decisions, short and some even long term plans on the basis of information we find on the Internet (What will the weather be like on my holiday? What are the market trends for the neighbourhood in which I am planning to buy a new family home? What food sources are healthy for me? What drugs are related to causing cancer?). The provenance of information (source, source credentials, trustworthiness and reliability, information dissemination path) coming from reliable and unreliable sources is hard, slow, and costly to verify. Also, the quality of the information in question is often uneven and unassessed. Someone with no credentials or expertise but with a large community gets high credibility on social networks and ultimately in mainstream media. Misinformation and malinformation get shared and propagate to unforeseeable extent. Given the right platform any information can appear as legitimate, and conflicts are often resolved unfairly. Even with a fair ontology (fair governance and recording process) information can be corrupted by malicious storage and network, or by censorship. With misinformation creating a new world disorder[5], the time for addressing data traceability and provenance is long overdue.

- ***Anonymity and unreliable identities:*** The practice of publishing anonymously or pseudo-anonymously has a long history in the arts, particularly in literature and journalistic or political writing. Even though there is no way to be truly anonymous on the Internet today there is a need to retain at least some amount of anonymity and protect the privacy of the people who need it. Fear of judgement, condemnation and retribution with the absence of identity protection will lead to a culture of fear and censorship, moving us away from the fundamental European values. Removing anonymity from the Internet should not be addressed as part of the effort to mitigate information disorder. Even with the anonymity removed, the issue of misinforma-

[5] https://www.scientificamerican.com/article/misinformation-has-created-a-new-world-disorder/

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

tion will remain: real people can and will provide false information for different reasons. Trustworthiness of different pseudo-identities should be properly assessed to mitigate misinformation. Moreover, to enable users to manage their own data and to facilitate anonymous trustworthy interactions, decentralized identity verification based on multiple attributes should be provided.

- ***No fair rewards for good quality contributions:*** Linked the remuneration issue is the problem of evaluating the quality of each single contribution (e.g. scientific paper, report, etc.). Various platforms publicly expose users' ratings as metadata over the public internet (e.g. rating of restaurants from Google, customer reviews for other goods and services like books from Goodreads.com), typically relating to the profile of single users. This model is flawed in two ways; first, it allows spam to mislead prospective consumers, while past consumers have little incentive in providing their feedback; second, the revenue that service providers make are not shared with the users that took the time to provide feedback. Beyond simple customer ratings and reviews, this problem applies to the reuse of users' contributions in all online services, and in social networks in particular. The main challenge here is to filter spam out in order to incentivize and reward quality contributions. On the ground of better content quality control, supporting quality reward systems brings together the concepts of truthfulness between multiple users on the one hand, and cryptocurrencies on the other. This feature, which is definitely lacking in the open Internet and that is a focus of NGI, will sparkle a new, fair ecosystem of better quality user-generated content.

- ***Bias in AI software:*** The under-representation of some social groups (Black, Asian and Minority Ethnic, people with handicap and victims of discrimination in general) both in privately owned companies and in governments de facto excludes those groups from contributing to ethical questions and discussions. For example, Amazon's recruiting engine was shelved because it was shown to unfairly discriminate against potential female hires. In 2015 Google's image search software identified a black software developer and a friend as gorillas[6].

  Identifying and mitigating bias in AI systems is essential to building trust between humans and machines that learn. As AI systems find, understand, and point out human inconsistencies in decision making, they could also reveal ways in which we are partial, parochial, and cognitively biased, leading us to adopt more impartial or egalitarian views. In the process of recognizing our bias and teaching machines about our common values, we may improve more than AI. We might just improve ourselves.

- ***Trustworthy blockchain service interoperability:*** Today there are hundreds of active blockchain projects in the GitHub repository . Dozens of new projects are emerging each year, competing with each other in the somewhat futile task of developing the best blockchain. Often, they would emphasize their product's alleged mar-

[6] https://www.theguardian.com/technology/2018/jan/12/google-racism-ban-gorilla-black-people

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

ket readiness, arguing that it is secure, scalable and overall better compared to a supposed rival. Regardless of whether their claimed characteristics are true or not, those projects represent stand-alone, disconnected blockchains. They entail different ecosystems, hashing algorithms, consensus models and communities. As a result, the blockchain space is becoming increasingly siloed, and its core philosophical concept –the idea of decentralization– is being undermined[7]. Focusing on trustworthy information exchange between multiple blockchains without an intermediary in the process, as well as integration with existing systems would allow them to be exploited to their full potential. This is the way to go, as according to Forrester's predictions for the DLT for the year 2020 interoperability is taking center stage[8].

2.1.2 Technology Challenge

The key point of ONTOCHAIN is to suitably federate blockchain and semantic technologies to overcome the aforementioned challenges. Doing this will require, among other things, facing the following technological calls:

- ***Decentralisation of heterogeneous components:*** ONTOCHAIN will leverage techniques, algorithms and software from many different fields (e.g. knowledge representation, storage and querying, Machine Learning, data analytics) and integrate them in a unique decentralized ontology framework. In addition to the effort of adapting their interfaces and semantics, porting each heterogeneous component to run efficiently and safely in a decentralized way, connected to a blockchain, will require specific adaptations.
- ***Fast pace of innovation in blockchains:*** the technologies that enable ONTOCHAIN are evolving so quickly, that many design choices will become obsolete before the end of the project. We are determined to make ONTOCHAIN sustainable by adopting a flexible approach, which translates into a set of guidelines for sub-projects. The most important guideline will be that each technical sub-projects produces two different results: a proof of concept that can be integrated into the ONTOCHAIN prototype, and generic design that can be reused outside of the project.
- ***Open design and flexible design:*** Sub-projects will have to make numerous trade-offs, e.g. between the granularity and how much data is stored on-chain vs. performance, that may evolve as future blockchain protocols emerge. Keeping these trade-offs documented and adaptable will help making ONTOCHAIN contributions interoperable and sustainable.
- ***Formal logic proofs:*** Another technological challenge posing itself is how to transparently derive a new truth out of several known truths according to a set of rules.

[7] https://cointelegraph.com/explained/blockchain-interoperability-explained
[8] https://go.forrester.com/blogs/predictions-2020-distributed-ledger-technology/

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

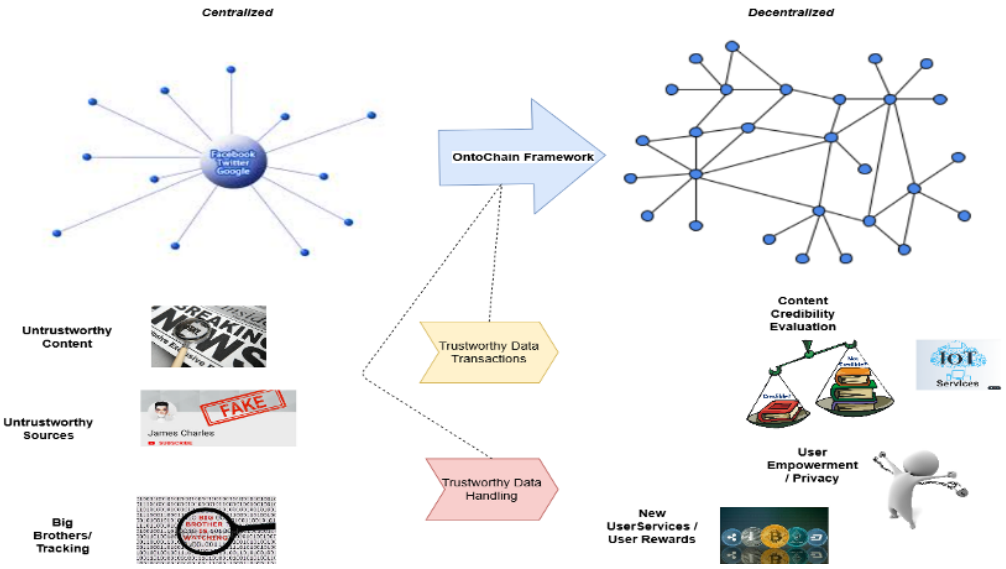

FIGURE 1: ILLUSTRATIVE VISION OF THE ONTOCHAIN PROJECT.

Various languages such as the OASIS Web Ontology Language (OWL) exist to express formal logic. OWL comes in three expressivity levels, known as species: Lite, Description Logic (DL) and Full. While formal rules are quick to compute in Lite, and more time consuming in DL, the Full specie cannot use formal proofs as the reasoning program may cycle indefinitely as proven by the already mentioned Kurt Gödel's theorems. There may, however, exist ways to design specific Smart Contracts that would implement first order logic directly on the blockchain

2.2 ONTOCHAIN OBJECTIVES

As shown in Figure 1, ONTOCHAIN envisions to shape a multi-layer and modular technology framework and to build on NGI and blockchain ecosystems and communities, 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 empower practitioners to address the various challenges of the Next Generation Internet through the use of multiple ledger technologies. ONTOCHAIN use-cases will be built upon different protocols and interactions between different blockchain frameworks, while hiding them from the use-cases to support effortless inter-service process cooperation. The proposed blockchain-based framework will enable higher performance and scalability, through the engagement of different business logics, access methods and governance models, whereas will present scalable solutions for ensuring secure and transparent content and information exchange as well as service in-

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

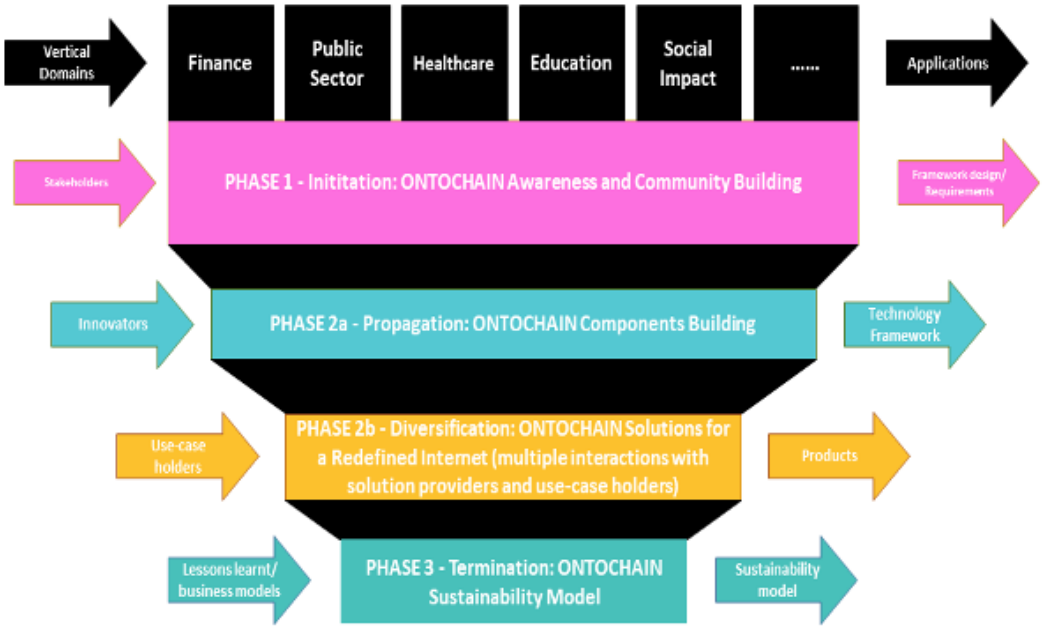

FIGURE 2: STRATEGIC PLAN OF ONTOCHAIN.

teroperability. ONTOCHAIN technology framework will constitute a building block of NGI towards a more human-centric Internet that supports values of openness, decentralisation, inclusiveness and protection of privacy and giving the control back to the end-users to be able to benefit from democratic, transparent and trustworthy decision making mechanisms.

ONTOCHAIN will reach its vision and strategic objectives by implementing an ambitious programme structured in four main phases (see Figure 2, modelling the steps of a chain reaction:

PHASE 1 – Initiation: To build a European ecosystem of stakeholders, that would include top researchers, innovators, technology developers focused on blockchain-based solutions, as well as representatives from different vertical sectors that are in the need of exploiting the ONTOCHAIN technology framework to cover their interests and business routines. The vision is to become part and contribute to the EU-blockchain community that will work together to design, produce, maintain the ONTOCHAIN concept and architecture, and as such become the pioneers to test and live the human-centric and trustworthy Next Generation Internet evolution. The representatives of this ecosystem will be invited to participate on the ONTOCHAIN materialization from Day 1, (a) setting the parameters for the development of the technical framework based on the real need as well as on the technologies that are available and from which the framework can benefit, (b) providing their expertise and technical know-how to build the framework,

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

and (c) adopting the proposed framework to serve their operations and address any challenges they face as end-users- towards a human-centric, decentralised and trustworthy web applications. In order the aforementioned to be achieved, an agile and effective funding mechanism will be designed and initiated, to act as the key enabler for all the stakeholders, as mentioned above, to be engaged, to research and develop important new ideas that contribute to the scope of the ONTOCHAIN project and the establishment of the Next Generation Internet.

PHASE 2a – Propagation: To create a technological framework to safeguard the Next Generation Internet from the direct or side effects of the aforementioned challenges; more specifically, ONTOCHAIN will work towards a more human-centric direction to the internet that will exploit state-of-the-art technologies (e.g. blockchain, ledgers, etc.) and novel concepts for the creation of a trustworthy information exchange process and a more transactional content handling. The ultimate goal is to re-invent the mechanisms of the Semantic Web on greater transparency basis, to move control from the large and centrally-structured corporations back to users, and at the same time to assure a trustful operation of services, data management, Quality of Service, GDPR compliance, etc. ONTOCHAIN aims at decentralizing that small but growing part of the Internet: web ontologies. As a central repository of knowledge and facts on which many applications rely, a web ontology is a critical unit of information storage and great care must be taken to ensure its content integrity, i.e. controlling who can record and modify ontological data. It has been recently shown that the way AI models are trained can lead to usage discrimination, in particular in applications like facial recognition for security, due to biased training sets that over-represented white faces compared to other ethnicities, rendering applications unusable for persons of colour. This example shows the danger of leaving facts like what is a face? in the hands of a single entity. In the end, public services relying on a centralized ontology will be more prone to censorship and discrimination. On the other hand, an ontology that is decentralized by design, i.e. that allows only to record facts based on a consensus of multiple agreeing entities (such as citizens) can benefit democracy by removing human bias from a large number of domains.

PHASE 2b – Diversification: To test real-life applications following the ONTOCHAIN agenda, assess the impact and calibrate the technical vision of ONTOCHAIN, for showcasing the business success of the ONTOCHAIN framework and its employed technologies for a trustworthy Internet. ONTOCHAIN will support a portfolio of test cases to implement the envisioned services of the ONTOCHAIN framework. These will be defined by the real needs of stakeholders, covering various vertical domains, e.g. eHealth, IoT, eScience, etc., that will be invited to join the ecosystem, contribute to the framework's formation and finally will be encouraged to bring their applications proposals and link them to the developed modules/ protocols of the framework, through a focused Open Call.

PHASE 3 – Termination: To create operations (sustainability) model to get a self-

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

sustained and organically growing ecosystem of actors, and allow for the deployment of ONTOCHAIN ecosystem's business based on the specific value proposition that it will offer to the stakeholders. The value proposition for Innovators (blockchain, cloud computing, software engineering experts) is the provision of a technical framework and close inter-relation with real-life applications, to build upon and access the wider market. This will simplify the business model of the participating research entities and will allow them to concentrate on their core technology proposition. The value proposition for the End-users (society-at-large) will be the offer of innovative customised and standardised applications for improving the operation and their relation within their audiences/ clientele. ONTOCHAIN will promote the developed framework as a systematic tool to be used by groups of people, communities, etc., across Europe in the implementation of the EU-Blockchain and NGI agendas and to influence internet governance and related policies. Within this same action the business models for long-term sustainability and expansion of best practices will be defined.

2.2.1 Specific Objectives

Having the above mentioned considerations in mind ONTOCHAIN is geared toward shaping a human-centric Internet (trustworthy, resilient, sustainable and inclusive) with the ability to provide semantic reasoning to the huge amount of data that each chain and contract generates through transactions. Moreover, ONTOCHAIN aims to form an integral, and pan-European blockchain ecosystem to address these challenges, to unleash the research and innovation potential of DLT enthusiasts across Europe, especially those in startup companies and innovative SMEs. The objective is to facilitate the take up of integrating research and innovation communities who focus on DLT across Europe by acting as facilitator, mentor and an enabler for those having both a clear idea and execution potential in line with the vision of ONTOCHAIN.

To measure reaching the intended goals and impacts, the following objectives have been defined:

Specific Objective 1 ONTOCHAIN ECOSYSTEM Setup: The blueprint ONTOCHAIN architecture, particularly its application and core protocols layers will be delivered as integrated and interoperable software and Application Programming Interfaces (APIs) under a commonly agreed open source licensing model, which is necessary in order to achieve interoperability of the solution. In addition to this, a scalable testing and production infrastructure will be formed allowing for seamless participation of various actors in the ONTOCHAIN ecosystem.

Specific Objective 2 ONTOCHAIN Technological Framework Design: The ONTOCHAIN technological framework will be designed in all its parts in order to address advanced use cases related to data provenance, decentralised reputation models, de-

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

centralised oracles, market mechanisms, ontology representation and management, privacy aware and secure data exchange, multi-source identity verification, value sharing and participation/contribution incentives and similar, and core protocols that include smart contracts, authorisation, certification, event gateways, identity management and identification, secure and privacy-aware decentralised storage, data semantics and semantic linking, ONTOCHAIN optimisation and similar.

Specific Objective 3 ONTOCHAIN Ecosystem experimentation: Experimentation will involve the introduction of innovative applications in many domains, including education, health, economy, mobility, public services, energy and sustainability, news, media, entertainment, Industry 4.0, tourism and so on. The project willl implement representative use cases in such domains.

Specific Objective 4 ONTOCHAIN Framework and Ecosystem Sustainability: Essential novelty of ONTOCHAIN are also its business models that apply to trusted knowledge intensive ecosystems of actors and resources. These will be designed to achieve a long-term sustainability of the ONTOCHAIN ecosystem. In the following subsection we also elaborate some key aspects in which the ONTOCHAIN Consortium believes it is possible to make substantial progress in near future.

2.2.2 Progress Beyond the State of the Art

In the following we elaborate just a few areas where the ONTOCHAIN Consortium believes substantial progress beyond the state of the art can be achieved.

Reputation management The development of generic decentralized reputation management functionality in the blockchain and different decentralized reputation models for trust assessment in various contexts are deemed as necessary components of the ONTOCHAIN ecosystem. Decentralized reputation mechanisms over blockchains promise to address several of the aforementioned issues, such as no single point of security/privacy vulnerability, stronger identities, privacy of the rating person, and more. There have been some initial instances of decentralized reputation systems on top of blockchain for various applications, as explained above and ONTOCHAIN will build on top of them to provide trustworthiness of subjects (people/content) without sacrificing user privacy.

Trusted data semantics ONTOCHAIN will bring data semantics in the blockchain framework, so as to track data provenance, data handling and data manipulations throughout the data lifecycle. Moreover, apart from data source trustworthiness, ONTOCHAIN will employ collaborative filtering techniques and data properties, to assess the truthfulness of data exchanged. Also, as compared to current approaches that employ unverified real-world data into smart contracts, ONTOCHAIN will verify the truthfulness of the data based on the concept decentralized oracles: arbitrary nodes that

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

are both distributed (highly available) and decentralized that approve the truthfulness of data through a consensus mechanism. iExec has already developed such a mechanism in the past[9], and within ONTOCHAIN, this mechanism can inspire the creation of reusable, complex decentralized oracles or similar concepts for several application domains that will be selected during the 3rd phase of the open calls of the project. Note that each application domain would require specific tuning of the oracles e.g. comparing/combining three pictures is very different from doing so with numbers. Another challenge for the success of a blockchain ecosystem is the ability to dynamically discover useful services based on user context and query semantics. ONTOCHAIN will allow the discovery of useful decentralized apps/services in the blockchain by means of semantic data annotation in the blockchain.

General data protection and marketplaces based on trusted knowledge and information ONTOCHAIN will address several challenges to unlock the tremendous potential of blockchains, especially before this paradigm shift becomes technically, economically and legally viable in business environments. The first category of these challenges concerns the technical aspects of blockchains including in terms of governance (i.e. open, private or consortium), scalability, data privacy, and validity of smart contracts. The set of challenges is related to the development of viable underlying business models and incentives mechanisms for user participation, for social welfare maximization, and so as the expansion of the system is economically sustainable. Last but not least, the legal aspects of blockchains represent a challenge, especially in France and Europe, where this technology should be analysed in the light of General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679) (EU 201657), whose objective is to strengthen users data privacy and protection within the European Union, and other related regulations.

[9] https://iex.ec/decentralized-oracles/

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

3 ONTOCHAIN TECHNICAL INVENTORY

This section introduces current open-source and free software projects that can be used as building blocks in ONTOCHAIN OC1. Software recommendations are made based on their innovation potential and their adoption by the corresponding communities in order to maximize the long-term impact of the software produced by ONTOCHAIN sub-grantees.

3.1 DISTRIBUTED LEDGERS AND SMART CONTRACT PLATFORMS

Ethereum which is the de-facto standard for smart contract based DLT applications, has been adopted by a large number of industrial and community projects. Ethereum is at the same time a production-level permissionless blockchain, and an ecosystem for building sidechains and Layer-2 chains. Ethereum blockchains are based on the Ethereum Virtual Machine (EVM) [1] for executing verifiable and auditable smart contracts that are primarily developed with the Solidity language. Ethereum Mainnet currently uses the Proof-of-Work consensus protocol for providing a very high level of trust in the chain's state and transactions. An important consideration is the upcoming Ethereum 2.0 and its shift towards Proof-of-Stake during the duration of ONTOCHAIN[10]. This shift, which aims at improving the performance of the chain while reducing its energy footprint will represent a major change but should not impact applications developed on top of the EVM.

Hyperledger Fabric (HLF) is an open-source framework for building, deploying and operating permissioned blockchains. Contrary to Ethereum, HLF blockchains support smart contracts developed in different programming languages including JavaScript and Java, allowing for cheaper and faster development. However, the limited performance of HLF chains when the number of validator nodes increases limits its applicability to small consortiums [2].

Tezos is a blockchain solution based on the Proof-of-Stake consensus protocol [3] that is intended to be easily upgradable and customizable through a built-in governance protocol involving a vote of the network's participants [4]. Tezos smart contracts are developed in the Lingo language, and a Python SDK is also available for compiling Python smart contracts to Ligo.

[10] https://ethereum.org/en/developers/docs/consensus-mechanisms/pos/

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

3.2 ONTOLOGICAL LANGUAGES AND DATABASESE

An ontology allows to define concepts and their relationship and properties in a subject area. Ontologies can be distinguished into Upper Ontologies, to represent general concepts, and Domain Ontologies, to formalize concepts in specific domains, as for example the domain of blockchain.

Different kind of languages have be used for representing ontologies, but the best practice which is consolidated since 20 years for the most popular ontologies is to use RDF and OWL, which are W3C standards and are related to each other. RDF[11] (*Resource Description Framework*) allows to make statements expressing relationships between resources, in the form `<subject> <predicate> <object>`; OWL[12] (*Web Ontology Language*) allows to obtain an higher level of expressiveness, by including the concept of classes.

Ontologies can be represented as graphs of concepts, relations and classes, and these graphs have to be stored in specific data bases. Different kind of implementation paradigms have been proposed for these data bases, hence his topic is substantially evolving. In summary, we can not specify a preferred data base or even a preferred paradigm to use: this is in fact a research topic for the ONTOCHAIN project.

Independently from the database used for storing ontologies, queries can be executed using the SPARQL[13] language, which is also a W3C standard. SPARQL is a recursive acronym for SPARQL Protocol and RDF Query Language.

In summary, in ONTOCHAIN we can specify the languages for representing ontologies (RDF and OWL) and a query language (SPARQL), not a data base to use, which is one of the research topics for Open Call 1.

3.3 DECENTRALIZED STORAGE

IPFS (the InterPlanetary File System[14]) is a Peer-to-Peer consensus and distributed hash table for storing and addressing arbitrary documents over computers connected by the Internet. On IPFs, files are immutable, addressed by their content and replicated over several nodes to ensure their availability.

FileCoin is a service and cryptocurrency built on top of IPFS, creating an economic and incentive model for users to contribute their free disk space. In FileCoin[15], IPFS

[11] https://www.w3.org/RDF/
[12] https://www.w3.org/TR/owl-features/
[13] https://www.w3.org/TR/rdf-sparql-query/
[14] https://ipfs.io/
[15] https://filecoin.io/

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

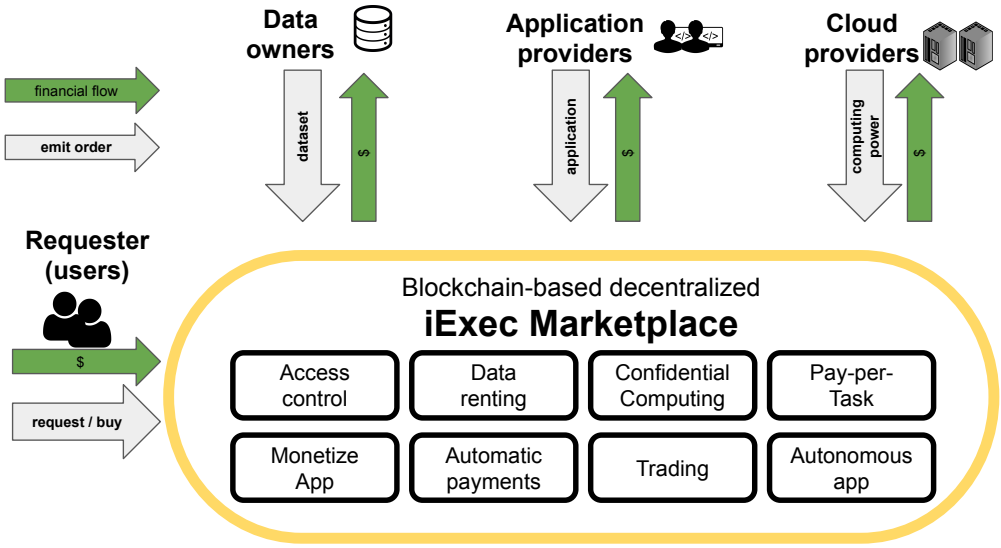

FIGURE 3: THE IEXEC MARKETPLACE

storage nodes have an incentive to keep storing a file as long as at least one agent in the network is still willing to pay for that file to be available.

Storj is an alternative to IPFS and FileCoin for storing files in a decentralized manner[16] over a network of nodes. Contrary to IPFS, files in Storj are encrypted by default and split over multiple nodes so that no node ever holds the complete file.

3.4 THE IEXEC PLATFORM

The iExec software connects cloud resource sellers with cloud resource buyers, encouraging an ecosystem of decentralized and autonomous, privacy-preserving applications. The platform has been developed with more than four years of efforts by iExec Blockchain Tech, a French company whose founders are former researchers from Inria and the Chinese Academy of Science. Since 2016, the platform has evolved through different versions, from V1 to V5.

The iExec network provides developers with scalable, secure and seamless access to decentralized services, datasets and computing resources. iExec's technology relies on Ethereum smart contracts and the Proof-of-Contribution (PoCo) protocol to create a transparent, resilient and secure virtual cloud infrastructure. Trust is ensured by incentive mechanisms, reputation and payments that are decentralized on the blockchain.

[16] https://www.storj.io/

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

Central to the network is the iExec Marketplace (see Figure 3, which offers a space where applications, datasets and computing resources (storage, CPU, GPU) can be monetized and provisioned uniformly, regardless of their provider. Because the Marketplace is decentralized on Ethereum, no one can control it: censorship is impossible and prices can only be determined by supply and demand.

Two types of users meet on the marketplace: resource providers (application providers, data providers, computing resource providers) and resource consumers (users, Web2 applications, smart contracts). The typical usage patterns for users from each group are described in the following sub-sections.

3.4.1 Usage patterns for Resource Providers

Application providers. Developers and application providers can monetize their apps, dapps, functions or algorithms. Just publish your application as a Docker container to the Marketplace and set a price per execution, it can now run on demand on multiple cloud providers.

Dataset providers. Data providers can monetize datasets usage and open up new revenue streams for their assets. Datasets must be published to the Marketplace just like applications, with a few extra steps. By encrypting their datasets, providers are guaranteed that they can only be decrypted inside the Secure Enclave within an Intel SGX processor, by an approved application.

Computing providers. Cloud providers can run the iExec worker on a set of physical or virtualized servers they wish to monetize; providers can either join an existing worker pool or create their own. Computing providers will be paid each time one of their workers perform a correct execution (per the Proof-of-Consensus protocol). By creating a worker pool and hosting a scheduler, they can get an additional share of revenue. Small providers can get access to a new revenue stream by monetizing some of their servers at times when they are underused.

3.4.2 Usage patterns for Resource Consumers

Users. People get access to a variety of cloud resources (CPUs, GPUs) from multiple providers to run applications without any vendor lock-in. Applications from the Marketplace as well as any other application are supported. By running their work on iExec, requesters can select higher trust, security and confidentiality, and get unique benefits:

- Trusted executions: the computation is replicated on several servers, and the result

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

is validated on-chain by reputation and majority voting in PoCo;

- Confidential execution: the data and the result are end-to-end encrypted, and the computation is run within a hardware enclave (Intel SGX); a cryptographic signature of the result proves its validity.

Web2 applications. Developers can bring the same benefits that people do by including the iExec SDK into their existing Web2 applications. Careful study of their application allows them to isolate its critical part and run it with higher trust, security and privacy on iExec with minimal change. Once integrated with PoCo, the application can render the same services as before and provide proofs of validity to their users and customers.

Smart Contract. With iExec, Smart Contracts can trigger an iExec task and get the result directly on-chain thanks to the callback mechanism embedded in PoCo. Smart Contracts can now query Oracles directly, and only when it is needed. On-chain validation rules can be set up to safeguard the smart contract even more and ensure that the resulting ingested satisfies its trust requirements.

3.4.3 Tasks execution and key features

The iExec platform supports tasks of two sorts Standard Tasks and TEE Tasks (i.e. Trusted Execution Environment). Standard Tasks are executed on untrusted resources and delegate trust to the PoCo smart contracts: a mix of replication, majority voting, economic incentives and reputation validates on-chain that the result is correct. TEE Tasks introduced in October 2018 add end-to-end encryption thanks to hardware cryptography. The tasks' data is only decrypted within an enclave and the result is encrypted and signed, which removes the need for replication.

3.4.4 Limitations and suggested improvements

This section covers limitations of the iExec platform that have been reported by the community and lists several improvements and integration projects that are deemed useful to the blockchain community at large. Applicants should note that the iExec platform relies heavily on Ethereum and is thus expected to benefit from the evolutions brought by Ethereum V2. Projects which leverage or integrate novel Ethereum features (e.g. Proof-of-Stake, Sharding, etc.) are encouraged. Listed here are some current limitations of the iExec platform along with suggested approaches:

- *Time-based payment* Currently, a task is considered valid only after it returns a verifiable result, meaning that services serving requests over an extended period of time

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

are difficult to port, and that the iExec worker cannot be paid based on how much time a task spent.

- *Multiple paid datasets per task* At the moment a task can only refer to one paid dataset (which does not prevent the application itself to download more data from the Internet). Strategies for allowing multiple datasets include hierarchical approaches or changes to the PoCo protocol.
- *Improved decentralized order brokering* While the verification process for sealing a deal (i.e. checking the validity of all the orders and the balance of the accounts) happens on-chain, the matching of orders (i.e. finding four valid orders and triggering the validation) is done off-chain. Matching orders is memory and compute intensive, thus a fully decentralized solution requires improvements of the algorithm and of the incentive model.
- *P2P order sharing* Currently, the essential mechanism for letting participants share and discover orders in the network is provided by a centralized repository. Because this repository sits outside the scope of the PoCo protocol it is not entirely censorship resistant: PoCo can validate that orders are valid but cannot check whether all orders are discoverable. A more decentralized approach could be to create a peer-to-peer order sharing network in which actors publish and search orders. Prototypes exist, but signal propagation related to the lifecycle of order is still not resolved in a secure manner.

In addition to this list, applicants are encouraged to find more possible improvements motivated by their own use-cases. Selected applicants will receive support from the iExec development team to design and implement their solution.

3.4.5 Current development agenda

This section covers possible developments that could benefit the Ethereum and the blockchain communities at large.

- *Complex Decentralized Oracles* In iExec, decentralized oracles can be arbitrarily complex and invoked directly from a client smart contract. This allows for the construction of a new generation of autonomous, decentralized oracles with self-enforcing governance rules. Generic oracles for performing aggregation between multiple data sources, multiple AI models for the same input data, sensor networks and so forth can have groundbreaking applications in virtually all domains.
- *Verifiers for ZK-Proofs* Zero-Knowledge proofs offer a promising opportunity for privacy and scalability in blockchains. Verifiers, however, hardly fit in a Smart Contract, meaning there is no onesize-fits-all solution for transferring trust from a ZK-Proof into a blockchain. This limitation is typically due to the size of the proofs (e.g. ZK-

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

STARKs) or to their execution time (e.g. ZK-SNARKs); implementing generic verifiers for several Zero Knowledge protocols in iExec tasks could let Smart Contracts verify arbitrarily proofs without worrying about their complexity.

- *State channel exit* Closing state channels usually require validating on-chain all the messages/transactions made in these channels as a necessary step to compute the outcomes. In most cases state channels are limited to simple cases (e.g. monetary transactions), because complex ruleset can become painful to process on-chain. Storing rulesets in iExec datasets and resolving state channels in iExec tasks could broaden the use of this technology to much more complex applications, by freeing the resolution process for the limitations of the EVM.
- *Marketplace of everything* Due to its initial purpose, the iExec platform is limited to selling and buying a fixed type of cloud computing resources. Supporting ontological representations of these resources could be a first step towards a much more generic platform in which new hardware can be instantly monetized. Eventually, a fully ontology-aware platform could monetize anything that can be modeled with a Semantic Web language, from servers to IoT devices, sensors, connected vehicles, 3D-printing models and so on.

3.4.6 Ecosystem integration

Suggest integrations between iExec and other projects (chains, web2 platforms, existing infrastructure projects etc.) that could provide valuable products to the community are sought by iExec.

- Wallets Support for new Ethereum wallets (including Smart Contract Wallets) brings support to more existing applications and is an important step towards end-users' adoption.
- Sensor networks Integration with emerging IoT protocols will eventually allow developers to provision connected objects along with servers, which is a critical enabler for Smart City applications and cyber-physical systems.
- Energy blockchains linking the iExec platform to Decentralized Energy Markets such as Energy Web paves the way to verifiable ecological and sustainable applications, autonomous green incentive systems and more.
- Kleros Integration with the Kleros Decentralized Court could permit applications to execute complex logic (e.g. compensations and penalties) in reaction to rulings;
- Cloud Controllers Drivers for cloud controllers, hypervisors and container engines such as OpenStack, Kubernetes or VMWare could lower the barrier to entering the virtual infrastructure and increase its size and heterogeneity. Smart drivers would

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

prioritize domestic workload and join the decentralized network when resources are underutilized, opening a new revenue stream to their owners.

3.4.7 Where to start?

iExec will support V5 of its software stack for the whole duration of the ONTOCHAIN project; candidates and selected participants are encouraged to refer to the Technical Documentation (https://docs.iex.ec/) and to the iExec White Paper [17] in order to familiarize themselves with the platform and consider if and how it can serve their project. The iExec software that can be used by ONTOCHAIN participants as a starting point to their evolutions and as a development platform for their applications is distributed under the Free Software license Apache v2.0; public source code repositories are hosted on the GitHub platform: https://github.com/iExecBlockchainComputing. In addition, please consider the following resources:

- iExec Academy, an aggregator for articles, tutorials, use-cases and media: `https://academy.iex.ec/`;
- The PoCo series articles explain the Proof-of-Contribution protocol and its evolutions: `https://medium.com/iex-ec/poco-series/home`;
- Example Dapps: `https://github.com/iExecBlockchainComputing/apps`;
- Dapps of the Week articles: `https://medium.com/iex-ec/dapp-of-the-week/home`
- Introduction to Confidential Computing: `https://docs.iex.ec/for-developers/confidential-computing/`
- How to Build a Data Privacy-Preserving App in Under 1 hour: `https://medium.com/iex-ec/how-to-build-a-data-privacy-preservingapp-in-under-1-hour-fb323e7458b`.

[17] `https://iex.ec/wpcontent/uploads/pdf/iExec-WPv3.0-English.pdf`

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

4 ONTOCHAIN FRAMEWORK & COMPONENTS SPECIFICATION

4.1 ONTOCHAIN ARCHITECTURE

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 4. This framework will enable the implementation of a number of different next-generation real-world solutions, such as trustworthy web and social media, trustworthy crowd-sensing, 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 logics, access methods and governance models that require specific chains. ONTOCHAIN use-cases will be built upon the different protocols shown in Figure 4. 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 (including data stored outside of ONTOCHAIN), 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. 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 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 in the text below, along with the dependencies among them.

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

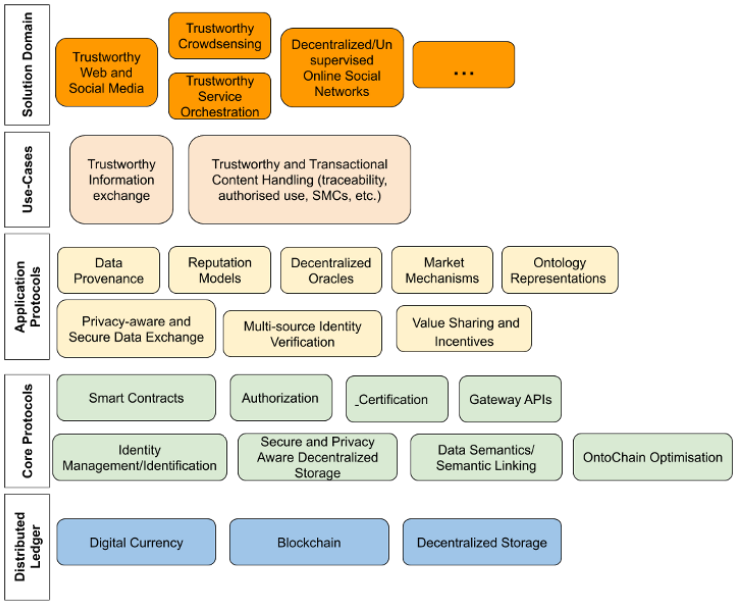

FIGURE 4: ONTOCHAIN ARCHITECTURE

4.1.1 Use-Case Layer

Trustworthy Information Exchange: This use case defines and develops 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 and transacted 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. Trustworthy and Transactional Content Handling addresses softer requirements for content handling where the decision 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 transfer among involved parties in a data transaction. Regulatory alignment of

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

data transactions, as a part of Trustworthy and Transactional Content Handling, will also seek to address the hard requirements for content handling. This means defining and developing 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. Regulatory alignment of data transactions seeks to apply regulatory rules, (e.g. requirement to process private data on SGX chips only), certifications for quality (e.g. only certified cloud providers allowed), specifically required point-to-point data transport protocols and mechanisms, or personal permissions (e.g. user allows to use their private data only in certain circumstances, such as an emergency medical situation).

4.1.2 Application Protocols Layer

Data Provenance: This module will provide graphical and programming interfaces for querying and presenting provenance information from ONTOCHAIN about on-chain and off-chain data (pointers to data stored outside of ONTOCHAIN). Provenance information will include the complete trail of transactions that resulted in a record, including links to the programs that were run (e.g. address of smart contracts, 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 infrastructure. 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. One important problem with reputation systems is weak identities, referred to as "cheap pseudonyms", through which multiple attacks can be employed, such as ballot stuffing, bad naming, negative discrimination, sybil attacks, reputation whitewashing, reputation milking and more. Existing solutions include reputation cold start (which introduces a new set of problems) and making pseudonym change more costly. Emerging decentralized reputation mechanisms built upon the Blockchain will enable stronger user identities without sacrificing anonymity. Different reputation models can be defined to assess different aspects, such as data source trustworthiness, data credibility, service trustworthiness, etc. This module is built upon Identity Verification mechanisms.

Decentralized Oracles: 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 could not be deterministic, and no consensus on their transactions could ever be reached. However, recording data from the real world into the chain is often necessary, and it will be a major requirement for ON-

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

TOCHAIN. The usual way of feeding data to a Smart Contract is through Oracles. An Oracle is a trusted off-chain program that can, at the request of a Smart Contract, examine real-world data and return it to the Smart Contract. Oracles are very critical in the sense that they create a point of centralization that goes against all Blockchain principles. Current approaches, such as the Quorum protocol, Substrate or ChainLink address the centralization issue by having multiple instances look at a data source, and then run a consensus algorithm onchain to validate the result. This, however, only displaces the point of centralization from the Oracle to the data source. Recent works on Decentralized Oracles propose to enforce that some data can only be recorded on-chain if several Oracles that report data from multiple data sources can reach a consensus. 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 (e.g. images, videos and text) can become complex. Solutions like iExec DOracles offer a platform for building Decentralized Oracles, but feeding real-world data to Smart Contracts with a satisfying level of trust will require innovative approaches that combine all components of ONTOCHAIN, including reputation, identity, Machine Learning and IoT.

Market Mechanisms as-a-Service: One of the grand purpose of blockchains is to support various market mechanisms. In fact, blockchains are prolific in this context, and one could imagine the formation of market mechanisms as-a-Service in the very near future. The trading of actual physical objects, but also of software services, data and information, nowadays happens through the Blockchains. This module provides the basic support mechanisms for enabling data/service transaction, and thus enables market mechanisms. For example, there are various exchanges that facilitate trading of assets and facilitate price determination (e.g., auctions, negotiation protocols, etc.), billing and customer support and more. It also provides functionality for enabling the sharing economy, such as value chaining, value/cost sharing and p2p cryptocurrency exchange. Moreover, it might include data value estimation approaches, algorithms for resource consumption estimation and associated costs and data predictions.

Secure Data Exchange: Secure data exchange 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 block seeks to define new ways for implementing ontologies with the use of Blockchain. Common components of ontologies include: (1) Individuals, such as instances, persons or objects (the basic or "ground level" objects, but they can be also abstract), (2) Classes, such as sets, collections, concepts, classes in programming, types of objects or kinds of things, (3) Attributes, such as aspects,

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

properties, features, characteristics or parameters that objects (and classes) can have, (4) Relations, such as ways in which classes and individuals can be related to one another, (5) Function terms, such as complex structures formed from certain relations that can be used in place of an individual term in a statement, (6) Restrictions, such as formally stated descriptions of what must be true in order for some assertion to be accepted as input, (7) Rules, such as statements in the form of an if-then (antecedent-consequent) sentence that describe the logical inferences that can be drawn from an assertion in a particular form, (8) Axioms, such as assertions (including rules) in a logical form that together comprise the overall theory that the ontology describes in its domain of application. This definition differs from that of "axioms" in generative grammar and formal logic. In those disciplines, axioms include only statements asserted as a priori knowledge. As used here, "axioms" also include the theory derived from axiomatic statements, (9) Events, such as the changing of attributes or relations, and at the pinnacle, (10) any reasoning approaches, tools and methods that can help deduce new knowledge arriving from a sensing IoT empowered environment.

Multi-source Physical/Abstract Object/Person Identity Registration and Verification: This block seeks to register and verify individual digital identities of physical objects (in addition, abstract objects, physical and abstract persons) via newly designed ONTOCHAIN services. Various AI methods could be introduced to operate on sensing data (IoT based, sensors, cameras and similar) so that an assertion can be made whether an individual belongs to a specific ontological concept (e.g. car, chair, container image, the person Elisabeth, and similar).

Value Sharing and Incentives: The 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 (which is common practice) and execution of smart contracts, ii) contribution assessment, e.g., Shapley values, etc. These are facilitated through an appropriate accounting system for measuring resource consumption for blockchain tasks.

4.1.3 Core Protocols Layer

Smart Contracts: Smart contracts are programs that are executed by several nodes of a blockchain (e.g. Tezos, Malboge, or Ethereum) that can directly read and write the state of the blockchain. Their output is an update of the blockchain's state and must be approved by the chain's consensus protocol. Smart contracts are *self-executing*, meaning that they can pose specific conditions under which a function can be executed, such as the triggering a method or a transaction. Some smart contract programming languages and environments allow the inclusion of oracles (e.g. iExec, Chainlink) that

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

makes it possible to make on-chain decisions based on information from the real world (i.e. outside of the chain).

Certification: Certification 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 cases, certificates can be issued on-chain, and can be used as conditions for performing specific transactions, for example, using AI methods to analyse private data. The 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: Secure Storage Solutions already exist on Blockchain. Together with decentralisation they help reduce the risk of having access to all private data. Moreover, various partitioning, fragmentation and redundancy methods are being used. An example of such a service is storj.io.

Identity Management: Self-sovereign Digital Identity is a specific area of research that is overreaching to be addressed solely by the present project. Nevertheless, ONTOCHAIN technologies and solutions can be used to address parts of the digital identity puzzle. There are two conflicting requirements that drive this development. First is the ability to identify oneself in specific interactions, such as withdrawing money in a bank, and another is to still preserve one's privacy, for example of the health data or the web browsing or buyer's habits. This is a feasible endeavour. However, it is necessary to invest more in technologies like ONTOCHAIN to make it happen.

ONTOCHAIN Optimisation: This module will provide new semantics-related solutions and will seek to minimise the necessary amount of both onchain and off-chain transactions, in order 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-chain transactions.

Gateway APIs: This module will support connections between the ONTOCHAIN Blockchain and the outside world, including other Blockchains. Part of its duty will be to help programmers in the upper layers make trade-offs about how much data is stored on-chain, by supporting pointers to offchain 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 at least three programming languages that are commonly used by developers e.g. JavaScript, Java and Python. The interfaces will be generic and extensible in order to allow connections with different ledger technologies in the future, while only external Ethereum-based chains will be supported during the course of the project because of its important community of adopters and developers and because its ecosystem already contains most of the

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

software components that sub-projects will require (e.g. off-chain Computing, Decentralized Oracles). However, sub- grantees will be discouraged from producing designs relying on concepts and optimisations that are specific to any particular Blockchain.

Data Semantics: Ontologies a core building block of the Semantic Web[18]. The W3C consortium provides mechanisms for their standardisation in order 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. Since ontology engineering is a complex work that usually takes many years to complete and test, the ONTOCHAIN project 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.

Authorisation: Blockchain has stimulated the idea of self-sovereign digital identity, and few commercial services have already emerged[19]. Various Role-Based Access Control (RBAC) systems have also existed in the course of the past decades. With ONTOCHAIN one could easily see systems where a patient is self-identified on the Blockchain, while a medical doctor gains access to the medical records based on their role (e.g. surgeon, general practitioner).

4.1.4 Distributed Ledger Layer

Blockchain Consensus Engine: Consensus making mechanisms are at the core of every Blockchain. ONTOCHAIN will be designed to be scalable, open, cost and energy-efficient, perhaps even elastic, it is necessary to design an 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.

Cryptocurrency: Beyond the current hype revolving around Bitcoin, Ethereum and over 5,000 altcoins, the potential for social change of what is now being called Blockchain 2.0 is appearing more and more clearly. For example, cryptocurrencies 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 selfexecuting transactions in Smart Contracts.

[18] https://www.w3.org/TR/?tag=data
[19] 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

Integrated with ONTOCHAIN's provenance and reputation mechanisms, a crypto token 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. In recent years, 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, for example, the Storj service. In the near future, one could imagine new storage services that can help store private data in encrypted and decentralised ways, that can help manage data replicas for reliability and Quality of Service, while balancing the trade-offs with the storage costs.

4.2 FRAMEWORK INTEGRATION

Different open-source solutions already exist for the design and integration of the different layers of the ONTOCHAIN architecture. Developers are encouraged to rely on the blockchain platform provided by iExec which already integrates low level layers of an analogue architecture i.e. application protocols, core protocols and distributed ledger. It supports public, private and federated blockchain solutions and connects cloud resource sellers with cloud resource buyers, encouraging an ecosystem of decentralized and autonomous, privacy-preserving applications. The iExec platform supports a decentralized marketplace of applications, data and resources for decentralized cloud computing in Ethereum.

In the case that they can not integrate their solution with the iExec platform, we suggest that funded third parties at least integrate it over the same mainnet the iExec platform, i.e Ethereum, or at least provide bridges so that their functionality is available to dapps hosted on Ethereum and to smart contracts running on the EVM.

In any case, the ONTOCHAIN project plans to (re)design, extend and otherwise integrate existing platforms and solutions, so that they can become integral and interoperable parts of the ONTOCHAIN software ecosystem.

Last but not least, the applicants are encouraged to use standards and widely used protocols to allow graceful integration with existing software. With this regard, a community approach is necessary to be able to advance DLT research and innovation and properly integrate open data and semantics with current and future blockchains.

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

5 CONCLUSION

This document summarizes the general challenges and objectives that ONTOCHAIN is tackling, and details the functional architecture of the project. An important concern of the deliverable is to provide third parties with recommendations when selecting technologies and building blocks for implementing the ONTOCHAIN architecture.

Applicants to ONTOCHAIN open calls for participation and selected third parties are also strongly encouraged to rely on the iExec platform whenever it is possible. In addition to integrating the lower layers of the ONTOCHAIN architecture on top of the Ethereum blockchain, iExec provides valuable functionalities including a decentralized marketplace of applications, data and computational resources which allows for transparent and traceable execution of applications and data processing over untrusted servers and virtual machines.

In the case when it is not possible to use the recommended open-source and free-software components, including the iExec platform, we strongly suggest that the provided solutions be deployed over the same mainnet (i.e Ethereum mainnet) or, in cases where the EVM cannot be supported, to at least provide bridges so that their functionality is available to dapps hosted on Ethereum and to smart contracts running on the EVM.

In any case, applicants are strongly encouraged to use standards and widely used protocols to allow for the graceful integration of their solutions with existing software.

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

REFERENCES

[1] Vitalik Buterin et al. *Ethereum white paper*. `https://ethereum.org/en/whitepaper/`. 2013.

[2] Tien Tuan Anh Dinh et al. "Blockbench: A framework for analyzing private blockchains". In: *Proceedings of the 2017 ACM International Conference on Management of Data*. 2017, pp. 1085–1100.

[3] Michael Neuder et al. "Selfish behavior in the tezos proof-of-stake protocol". In: *arXiv preprint arXiv:1912.02954* (2019).

[4] L.M. Goodman. *Tezos white paper*. `https://tezos.com/whitepaper.pdf`. 2014.

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