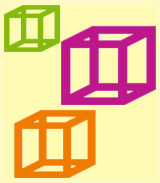
A more pragmatic Web 3.0: Linked Blockchain Data

INFORMAL KHAKI RESEARCH PAPER



Msc. Héctor E. Ugarte R.
Rheinische Friedrich-Wilhelms-Universität Bonn - Germany
March 18, 2017

During: Bonn - Cusco - Lima |>-

***Abstract*—Linked Data is proclaimed as the Semantic Web done right. The Semantic Web is an incomplete dream so far, but a homogeneous revolutionary platform as a network of Blockchains could be the solution to this not optimal reality. This research paper introduces some initial hints and ideas about how a futuristic Internet that might be composed and powered by Blockchains networks would be constructed and designed to interconnect data and meaning, thus allow reasoning. An industrial application where Blockchain and Linked Data fits perfectly as a Supply Chain management system is also researched.**

Keywords: Blockchain, Linked Data, Semantic Web, Semantic Blockchain, BLONDiE, Supply Chain, Bitcoin, Ethereum, IPFS.

I INTRODUCTION

Blockchain (a.k.a. Distributed Ledger) is a new paradigm being used lately as the core of different decentralized crypto-currencies platforms and as a trend on the FinTech (Financial Technology) industry. The original and starting revolutionary project is Bitcoin[1]. While Bitcoin is used widely as a successful decentralized financial transaction system, other proposed platforms are trying to cover other use cases more than just the crypto-currency. These platforms will enable plenty third-party projects that among other things require the use of standards for data exchange.

Tim Berners-Lee (the creator of the world wide web) anticipated on the evolution from a Web of documents to a Web of data. The resulting technology is called "The Semantic Web", where a set of standards for data exchange are defined. However, some researchers claim that is impractical and theoretically impossible to totally achieve what it proposes. In fact, to have all the current Internet content semantically linked is a utopia, where many other technologies to handle Big Data using Data Mining and Machine Learning techniques exist. Semantic Web is widely used and popular on the academic world, but very misunderstood and underestimated on the business environment, even in 2006, Berners-Lee and colleagues stated: "This simple idea remains largely unrealized" [1].

Nevertheless since Blockchain is a new paradigm that can serve as the core framework of the Web 3.0, it is realistic to think that we can share all or part of its data from different levels of the protocol using some technologies already proposed by the Semantic Web community. Furthermore, it also results important to adopt techniques and existing semantic principles for the early architecture design of more mature futuristic Semantic Blockchains. These new networks, powered by Smart Contracts could be the core technology for artificial intelligent agents with reasoning ability.

After reading this paper, replace "Semantic Web" with "Blockchain" in the next quote and you will realize that it fits perfectly:

> *"...I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web: the content, links, and transactions between people and computers. A 'Semantic Web', which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The 'intelligent agents' people have touted for ages will finally materialize..."*
>
> — Tim Berners-Lee

II BACKGROUND

II-A *Blockchain*

A Blockchain is just a data structure, that can increase its size and is shared by different clients thanks to the peer-to-peer (P2P) architecture. But, since cryptographic tools are used, it is not possible to modify some data without the proper private keys. Even more, any modification will be stored in the chain and be public forever. It is structured as a chain of blocks (linked list), where each block has a set of transactions that occurred during a specific period (time-stamped), As can be seen in Figure 1.

*This work was not supported by any organization
Twitter: https://twitter.com/hectugaroj
More work on the Semantic Blockchain:
https://semanticblocks.wordpress.com/

[1]https://www.bitcoin.org/

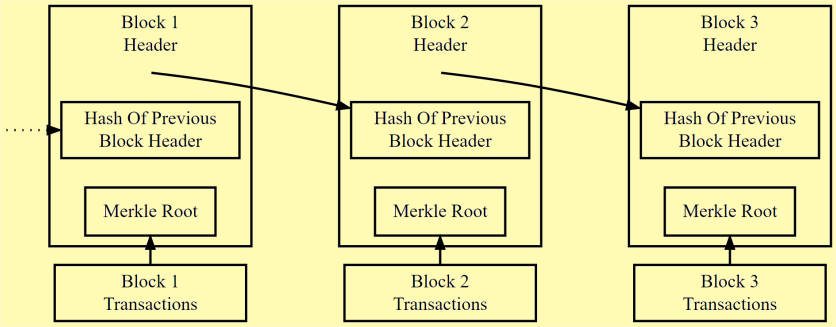


Fig. 1. Simplified bitcoin Blockchain structure [2]

Definition 1. Blockchain. *The longest path from the genesis block (root of a tree) to the leaf is called Blockchain. The Blockchain acts as a consistent transaction history on which all nodes eventually agree* [3].

Definition 2. Block. *A block is a data structure used to communicate incremental changes to the local state of a node. It consists of a list of transactions, a reference to a previous block and a nonce* [3].

Definition 3. Transaction (bitcoin). *A transaction is a data structure that describes the transfer of bitcoins from spenders to recipients* [3].

Some of the benefits and unique features of this new paradigm are:

- **Ownership of data.** Ownership of digital, physical records and assets of any kind can be proved.
- **Uniqueness and proof of uniqueness.** The uniqueness of a digital, physical record and asset is guaranteed.
- **Immutability.** Some data stored on a Blockchain is accepted as valid, and cannot be modified.
- **Censorship resilient.** The existence of censorship like firewalls blocking access to data is not possible.
- **Public transparency and traceability.** Anyone can see the content of the transactions and audit them.
- **Trust-less and incorruptible.** Blockchain allow us to build trust-less systems using P2P, such that contracts can be encoded without trusted third parties.
- **Cost-efficient.** Costs are reduced because there are no external actors and also because there is a unique homogeneous platform for different tasks.
- **Guaranteed continuity.** The continuity of the system is guaranteed as far there are nodes running the chain.

These features allow the Blockchain to be a potential path for artificial intelligence (AI) [4], fundamental for an optimal realization of the "Internet of Things" and other Industry 4.0 technologies. In the journey emerge a lot of research to be done in the convergence of machine learning and distributed ledgers, For instance, a scenario using Deep Learning and Neural Networks powered by a distributed ledger.

II-B Smart Contracts

Nick Szabo coined the term Smart Contract to define a tool to automate human interactions [5]. Only since the appearance of Bitcoin and not before, there exists a platform to program them as an algorithm that can self-execute, self-enforce, self-verify, and self-constraint the performance of the contract [6, p.16].

Listing 1. Ethereum Smart Contract Example [7]

```
contract mortal {
    address owner;
    function mortal() {
          owner = msg.sender;     }
    /* Function to recover the funds on the
       contract */
    function kill() {
          if (msg.sender == owner)
            suicide(owner);     }
}
```

Bitcoin per se is the first Smart Contract ever developed. A platform to code, execute and share this kind of contracts as Turing-complete programs is Ethereum[2] (example in Listing 1). Hyperledger[3], a permissioned business oriented Blockchain framework, calls them as "Chain code". Frameworks like these two and many others enable the possibility to build decentralized applications (ĐApps) powered by Smart Contracts.

II-C Linked Data

According to Tim Berners-Lee, Linked Data is the "the Semantic Web done right, and the Web done right". When information is presented as Linked Data, other related information can be easily discovered and new information can be easily linked to it. It is based on these 4 rules [8]:

i) Use URIs (Uniform Resource Identifiers) as names for things.
ii) Use HTTP URIs so that people can look up those names.
iii) When someone looks up a URI, provide useful information, using the standards. (RDF, SPARQL)

[2]https://www.ethereum.org/
[3]https://www.hyperledger.org/

iv) Include links to other URIs. so that they can discover more things.

II-C1 RDF

The Resource Description Framework (RDF) is a family of W3C specifications, "it is a foundation for processing metadata" [9]. On web resources, RDF is used as the standard way to describe and model information. Three object types conform the basic model [9]:

i) **Resources.** The things that where RDF expressions are used to describe them.
ii) **Properties.** The specific description of a resource, it can be an attribute or a relation.
iii) **Statements.** The conjunction of a resource, a named property and the value of that property. These three elements form the RDF statement of a specific resource. They are expressed in the form of subject, predicate, object and commonly called "triples". Triples create a basic graph structure of data.

Definition 4. RDF triple. *An RDF triple t is defined as a triple t = (s,p,o) where* $s \in U \cup B$ *is called the subject,* $p \in U$ *is called the predicate and* $o \in U \cup B \cup L$ *is called the object. where: U (Set of all URIs), L (Set of all literals) and B (Set of all blank nodes)* [10].

II-C2 SPARQL

SPARQL (recursive acronym for SPARQL Protocol and RDF Query Language), is a W3C recommended semantic query language for datasets, made for retrieving and handling data stored in RDF format. Thus, the queries are working over a graph structure defined by the RDF data, where the result will also be a graph or a subset of it.

Definition 5. SPARQL dataset. *A SPARQL dataset* $D = \{G_D, (u_1, G_1), ..., (u_n, G_n)\}$ *is a set of graphs where* $u_1 ... u_n$ *are distinct URIs and* $G_D, G_1, .., G_n$ *are RDF graphs.* [10]

II-C3 Ontologies and OWL

An ontology is a set of explicit formal specifications of the terms (classes or concepts) in a domain and relations (properties or roles) between them [11]. When a set of individuals (instances) is available, it is known as a knowledge base. Ontologies define a common vocabulary for researchers that want to share information in a domain and they include machine-readable definitions of basic concepts and its relations [12].

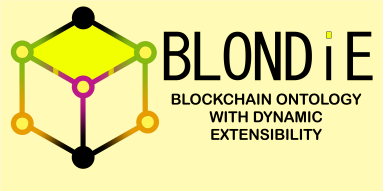
Web Ontology Language (OWL) is a language made to represent complex and rich knowledge about things, sets of things and existing relations between them. OWL documents or OWL ontologies are usually published on the WWW and make reference or be referred from others OWL documents. OWL is a computational logic-based language meaning that knowledge modeled in it can be exploited by computer programs, e.g. to make implicit knowledge explicit available [13]. It has a rich set of operations like union, negation, intersection, etc. Its logical model allows the use of reasoners that are checkers of consistency between ontology elements.

II-D Web 3.0

The evolution and interaction of people on the Internet emerged on a classification of this revolutionary technology. Early websites formed what is now known as Web 1.0 the "web of documents", documents serving as information portals with basic ability to link websites. Later the Web 2.0 emerged as the "web of people", it introduced users to collaborate with content creation and alteration. For the coming Web 3.0 there is a debate about what is the proper definition of its characteristics. For some group, the Web 3.0 is powered by the Semantic Web, where people can access to linked information fast and easy. But now, with the emergence of a decentralized web powered by Blockchain technology and since it enables unmediated transactions, there is a new focus on the Web 3.0 based on through the trustful nature of the blockchain. It is the "read-write-own web". Here, the user owns and participate in owning the protocol. It is both peer to peer and machine to machine. And it is applicable to people, companies and autonomous entities [14].

For instance, the term Web 3.0 is used by Ethereum in a different context than the suggested by Berners-Lee. It is proposed as the separation of content from the presentation by removing the need to have servers at all [15]. Stephen Tual, Ethereum's CCO's, defines that makes Ethereum different than Web 2.0 is that "there are no web-servers, and therefore no middleman to take commissions, steal your data or offer it to the NSA, and of course nothing to DDoS" [16].

III BLONDiE: Blockchain Ontology with Dynamic Extensibility



An initial question to answer is: Why to develop an ontology?. According to Noy and McGuiness [12], some of the most relevant reasons are:

- To share common understanding of the structure of information among people or software agents.
- To enable reuse of domain knowledge.
- To make domain assumptions explicit.
- To separate domain knowledge from the operational knowledge.
- To analyse domain knowledge.

BLONDiE is an OWL ontology for describing the Blockchain native structure and related information. In its current version (0.4), it covers the two most relevant cryptocurrencies in the moment: Bitcoin and Ethereum. As an initial step, for instance, if we represent data from Bitcoin and Ethereum with RDF, it will be possible to link a person with his accounts (Figure 2).

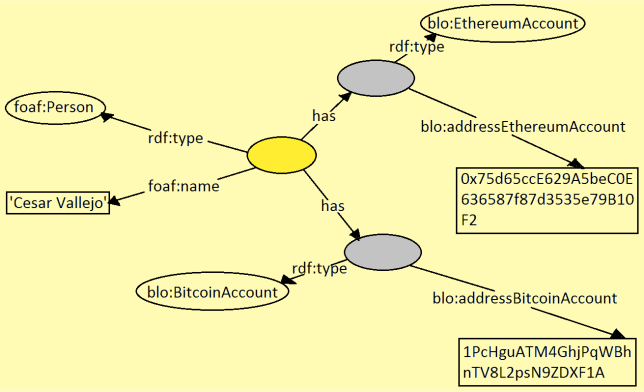


Fig. 2. FOAF and BLONDiE usage example

III-A Representational Requirements

There are two main goals that we want to achieve: First, to developed a schema for a queryable knowledge base that stores information from the Bitcoin and Ethereum Blockchains native structure and other related information and second, to include business intelligence in the knowledge repository that runs on powerful semantics to answer queries from the users about the Bitcoin and Ethereum distributed ledgers.

There is a natural need to browse the content of existing Blockchain frameworks like Bitcoin and Ethereum, web applications with this ability are known as Blockchain browsers, most of these existing tools use relational databases or key-value databases and not graph databases fully compatible with a machine readable format as RDF.

Some competence questions that our knowledge base should answer are:

- **CQ1.** Who was the miner of each block?
- **CQ2.** What is the height of each block?
- **CQ3.** How many transactions were included in a block?
- **CQ4.** Is a transaction confirmed or unconfirmed?
- **CQ5.** How many total coins were transferred on a block?

III-B Domain Capture

An overview of relevant conceptual entities and types of relationships is done in this section. The structural information of Bitcoin and Ethereum Blockchains and related classes as they are expressed in the official documentations [17] [18] is gathered. Later we collect other information according to our previous analysis. Our domain capture is presented in Figure 3 in the form of an Extended Entity-Relationship diagram.

III-C Result and other ontologies

The result of the development is the BLONDiE ontology formatted on the file blondie.owl available on the Github repository[4]. In BLONDiE vocabulary we defined 21 classes, 11 object properties and 50 datatype properties.

[4]https://github.com/hedugaro/BLONDiE/

On Early 2014, a draft formally unnamed ontology[5] was published by Melvin Carvalho. It is made to be functional for cryptocurrencies like Bitcoin, Bitmark, Ripple, altcoins and others, but for its simplicity, limitations and not consider that each cryptocurrency can have different properties and classes, it does not seem to be very practical and functional for representing completely structural data from the Blockchain. This ontology uses RDFS as data modelling language for describing RDF data. The published ontology is described as being in development. It describes 7 classes and 31 properties. None of the properties have domains or ranges specified in RDFS, it is lacking in a formal documentation and no real application using it is known.

BLONDiE is made to be extensible for other ontologies, so, integration to other works is easy. Recently a more specific Ethereum ontology called EthOn[6] was developed. It is defined as an Ethereum ontology that is closely aligned with the Ethereum yellow paper [18]. It is a very interesting work that among other use cases allows to semantically annotate content provided by Ethereum based tools and ĐApps.

There exist other works more oriented to the financial industries that in some cases consider some properties alligned to the distributed ledgers. FIBO (Financial Industry Business Ontology)[7] is a core ontology for the Financial Services domain, FIBO has more than 600 classes, some of them are relevant for the Blockchain. It can be aligned and matched with Ontologies like EthOn, finding correspondence between the EthOn concepts and FIBO clusters. In that way, a bridge that facilitates analysis and system design is created [19].

IV Linking Blockchains

On a general point of view, the Web is a huge global graph of connected hypertext documents by hyperlinks. On a similar way, there are some projects working in the idea of connecting different Blockchains. Where the main goal is to create “the Internet of Blockchains”. Since payments are different from plane information, it can be copied and replicated, but money must not be.

The interledger protocol (ILP)[8] is for payments across payment systems. ILP models the world of finance as a giant global graph of ledgers connected by liquidity. The systems providing this liquidity are called connectors and a key feature of ILP is that these connectors do not need to be trusted, meaning anyone can create one [21].

Currently, centralized exchange platforms are used to exchange one Blockchain-based currency for another. Cosmos[9] is a network of Blockchains, organized on hubs and zones. Zones plugged into a central hub and each zone maintain its each governance. Allowing the decentralized exchange of tokens from one Blockchain to another [22].

[5]https://github.com/melvincarvalho/crypto-currency-ontologies
[6]https://github.com/ConsenSys/EthOn
[7]http://www.edmcouncil.org/financialbusiness
[8]https://interledger.org/
[9]https://cosmos.network/

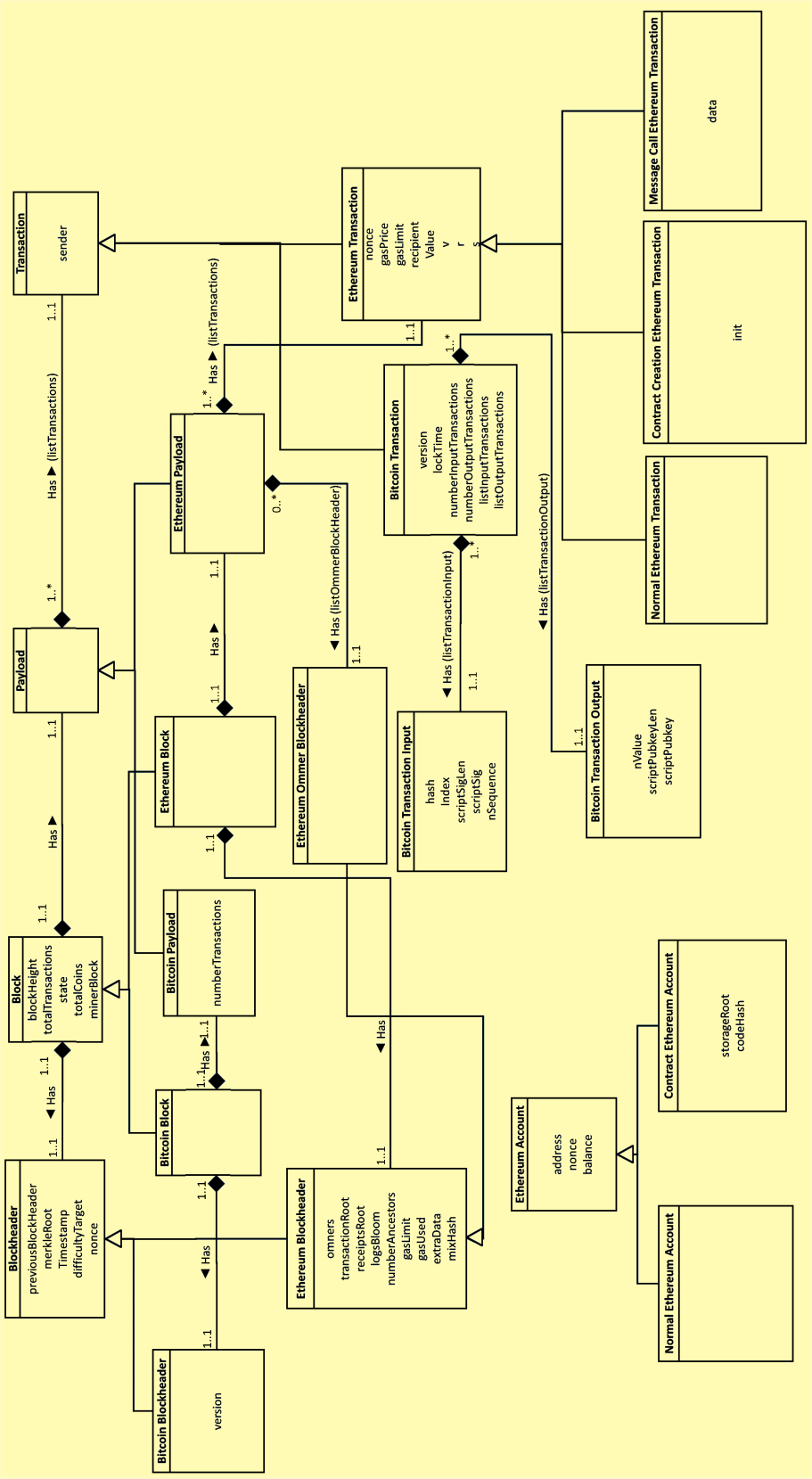


Fig. 3. Extended Entity-Relationship Diagram of BLONDiE

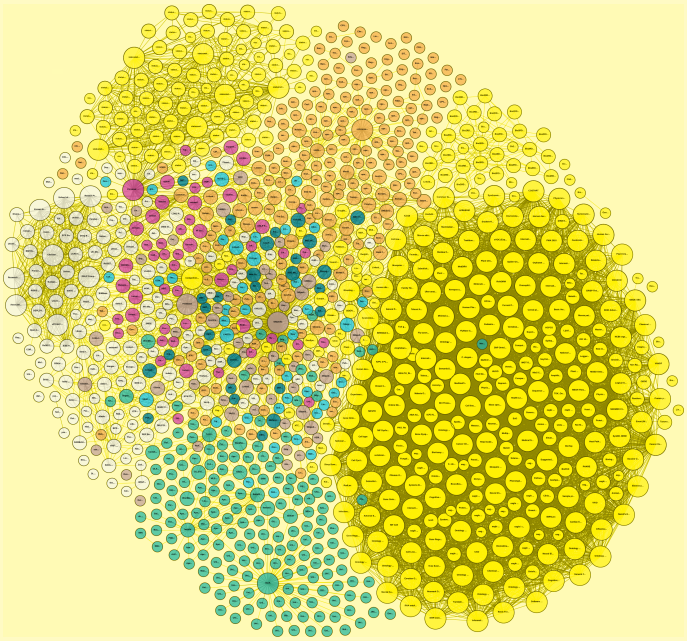


Fig. 4. The Linking Open Data cloud diagram [20]

Polkadot[10] is a new network that aims to provide interoperability between private and public Blockchains. Allowing extensibility and scalability. It defines a heterogeneous multi-chain, provided to an absolute minimum of security and transport. Scalability is addressed through a divide-and-conquer approach. The heterogeneous nature of this architecture enables many highly divergent types of consensus systems interoperating in a trustless, fully decentralized "federation", enabling open and closed networks to have trust-free access to each other [23].

Two more similar projects are Blocknet[11] and Supernet[12]. All these projects show evident need of semantic empowerment. The same need that the web still has to really jump from an only "Syntactic Web" (Web of documents) to a "Semantic Web" (Web of data). While how and where to locate and enable semantic data is not completely clear, it is obvious that Semantic Web fits as a prominent set of technologies to be used here. In Figure 4, we present datasets that have been published in Linked Data format by contributors to the Linking Open Data community project and other individuals and organizations. In a similar way, Blockchains-based systems and its corresponding datasets can be published and linked.

V Semantic Smart Contracts

V-A *Ontology-Based Contracts*

Data models are used on Ontology-Based Object-Oriented Enterprise Modelling [24]. In a similar manner, Kim and Laskowski [25] propose that ontologies can contribute to Blockchain design. Their approach can aid in the development of Smart Contracts that execute on the Blockchain.

They come up with a proof-of-concept using Ethereum and TOVE Traceability Ontology. This interesting idea can be replicated for other ontologies too. Better data standards, business practices and processes for developing and operating Blockchain are achieved. It also aids in the formal specification for automated inference and verification in the operation of a Blockchain [25].

V-B *Ricardian Contracts*

The Ricardian contract is a method of recording a document as a contract at law and linking it securely to other systems such as accounting for the contract as an issuance of value [26]. It holds three properties:

i) **robust:** use of identification by Cryptographic hash function.
ii) **transparent:** use of readable text for legal prose.
iii) **efficient:** use of mark-up language to extract essential information.

"A Ricardian Contract can be defined as a single document that is a) a contract offered by an issuer to holders, b) for a valuable right held by holders, and managed by the issuer, c) easily readable by people (like a contract on paper), d) readable by programs (parsable like a database), e) digitally signed, f) carrying the keys and server information, and g) allied with a unique and secure identifier" [26].

The chosen format for the text of the contract can use Semantic Web valid encoding methods as JSON-LD, where how to improve the readability for average users has to be considered and researched. There is a natural intersection between Ricardian Contracts and Smart Contracts. A Smart Contract can execute a Ricardian Contract. "The Smart Contract is really the machine to perform the contract" [27].

Listing 2. OpenBazaar's Ricardian Contract structure [28]

```
1. Contract terms
    - Seller's nym
    - Contract nonce (unique identifier)
    - Seller's bitcoin pubkey (for multisig
       transactions)
    - Merchant Data
        - What is to be sold
        - Price per unit
        - Additional conditional detail
            unforseen in this proposal
    - Contract expiration date
    - Seller's PGP public key
2. All of the above is digitally signed with
   the seller's private PGP key
```

There are several explicit extant implementations of the Ricardian contract design pattern and there are some projects which do so implicitly and others which are heading in that direction. OpenBazaar[13] is a project developing a protocol for e-commerce transactions in a fully decentralized marketplace

[10]http://polkadot.io/
[11]http://blocknet.co/
[12]https://supernet.org/
[13]https://openbazaar.org/

using Bitcoin. OpenBazaar uses Ricardian Contract as a means of tracking the liability of one party to another when selling goods to each other. The structure is presented in Listing 2.

V-C Trust Contracts

Ethereum's Smart Contracts are based on Turing-complete language (able to answer computable problem given enough time and space). Using a fees mechanism called "Gas" is guaranteed that it will not run forever, but some researchers claim that this can bring some problems. Park et al. [29], propose a new system called Boscoin[14]. They instead suggest the use of "Trust Contracts", since Turing-Complete ones are inherently undecidable and difficult to be read by non-technical people, resulting not being easy to know what a contract will do before running it.

Trust contracts are based on OWL and TAL (Timed Automata Language). OWL provides the structure of data and TAL acts as an operator. These contracts are decidable, easy to read and is possible to determine the amount of time they take to run [29].

VI Decentralized Storage and RDF

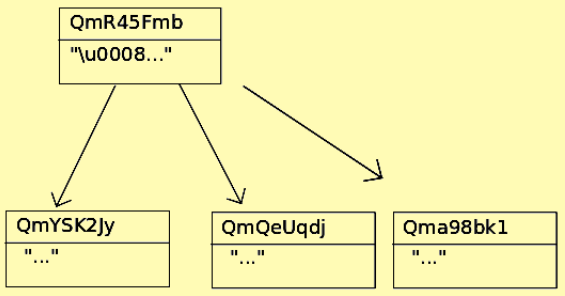


Fig. 5. IPFS graph on big files [30]

According to the W3C best practices for data on the web [31], it is important to "make data available on machine-readable and processable format on the web", This allows the creation and combination of data sets by users, and the capability to interpret data available in suitable formats and applications [32, p.4], just as the same way that is useful to make documents human readable available. For data the lingua franca is RDF [33].

JSON-RPC is a remote procedure call protocol that is encoded in JSON format. This simple protocol defines a few set of data types and commands. It allows sending data to a server that does not require a response and multiple calls which may be answered out of order. Geth, the most popular Ethereum client, offers a JSON-RPC endpoint.

Using the JSON-RPC command: eth_getBlockByHash, it is possible to verify the existing structure of a block. (See Listing 3).

[14]https://www.boscoin.io/

TABLE I
BITCOIN DATA STORAGE [35]

Method	Description
Value	Encode data in the number of satoshis being sent to an address.
Fake Addresses	Encode data in the Address itself. Because the Address encodes data of,your choice it cannot have been the result of a derivation from a,private key (with extremely high probability) and thus any coins sent to,such addresses are lost (or "burnt").
Vanity Addresses	Brute force through keys until you get an address that encodes your,data, extremely resource intensive and impractical for anything bigger,than a couple of bytes.
1 of N Multisig Address	These are more complex Bitcoin addresses that require one key out of N,to redeem. We can use only one key as a real key (like with a standard,address) and encode 32 bytes of data in the remaining N-1 keys.
OP_RETURN	OP_RETURN is a command in the Bitcoin scripting language that was,specifically added to allow the inclusion of metadata on the Blockchain.,Currently 80 bytes of information can be added to a transaction using,OP_RETURN.
Input Sequence	This is an unused 32 bit integer.
Coinbase	Miners can include up to 100 bytes worth of data in a coinbase transaction.

The Blockchain can also be used as a RDF repository. Even though it is not clear if this approach is the most optimal. Limitations on speed and size of machine-readable files need to be researched. On top of the Bitcoin Blockchain, there are several ways to store data, they are summarized in Table I.

On Ethereum, because it has a special structure and existing properties such as paying fees with Gas, it results relevant to research different possible ways to store data in general, and in particular RDF triples. In Figure 6 we present the framework elements. Most wallets use JSON as format to store data, this data can easily be converted to RDF. Also Decentralized Application's Front-end is generally made with HTML technologies, where we can use existing formats to embed data on the website interface as Microformats, RDFa, HTML5 Microdata, JSON-LD, etc.

The most relevant part of this research is the storage of data on the Blockchain itself. The only property that a block header offers to store other data different than the intrinsically needed is the "extraData" property. It allows us to store data of 32 byte size and this is defined by the winner mining node of this block.

The different methods to store data in Ethereum transactions are summarized in Table II. Finally, since there exist limitations of storage and fees to pay in the Ethereum Blockchain, it is desirable to use a compact RDF serialization format (for example Header-Dictionary-Triples). Different proposals and studies covering it exist, but since that is not the aim of research of this paper, we will not focus on that aspect.

Listing 3. Ethereum Block Structure Example [34]

```
{
"id":1,
"jsonrpc":"2.0",
"result": {
    "number": "0x1b4", // 436
    "hash": "0xe670ec64341771606e55d6b4ca35a1a6b75ee3d5145a99d05921026d1527331",
    "parentHash": "0x9646252be9520f6e71339a8df9c55e4d7619deeb018d2a3f2d21fc165dde5eb5",
    "nonce": "0xe04d296d2460cfb8472af2c5fd05b5a214109c25688d3704aed5484f9a7792f2",
    "sha3Uncles": "0x1dcc4de8dec75d7aab85b567b6ccd41ad312451b948a7413f0a142fd40d49347",
    "logsBloom": "0xe670ec64341771606e55d6b4ca35a1a6b75ee3d5145a99d05921026d1527331",
    "transactionsRoot": "0x56e81f171bcc55a6ff8345e692c0f86e5b48e01b996cadc001622fb5e363b421",
    "stateRoot": "0xd5855eb08b3387c0af375e9cdb6acfc05eb8f519e419b874b6ff2ffda7ed1dff",
    "miner": "0x4e65fda2159562a496f9f3522f89122a3088497a",
    "difficulty": "0x027f07", // 163591
    "totalDifficulty":  "0x027f07", // 163591
    "extraData": "0x0000000000000000000000000000000000000000000000000000000000000000",
    "size":  "0x027f07", // 163591
    "gasLimit": "0x9f759", // 653145
    "gasUsed": "0x9f759", // 653145
    "timestamp": "0x54e34e8e" // 1424182926
    "transactions": [{...},{ ... }]
    "uncles": ["0x1606e5...", "0xd5145a9..."]
  }
}
```

TABLE II
STORING RDF DATA ON ETHEREUM TRANSACTIONS SUMMARY [36]

Way	Short explanation	Advantages	Disadvantages
Transaction Data Property	Property existing in each transaction on Ethereum	- Not fixed size - Cannot be modified	- Expensive - Stored on hexadecimal format - Is not SPV friendly
Contract Storage	Contract state flexible database. Key-value store	- Not fixed size - Easily accesible	- Expensive - Information is modifiable
Event Logs	Historical raw data	- Cheap	- Not accesible for smart contracts. - Data generated by the Smart Contract
External Storage (IPFS)	Storing it externally and keeping the identifier using one of the above methods	- Unlimited size	- Not guaranteed that data will not be removed

VI-1 IPFS

The InterPlanetary File System (IPFS) is an open source project initially designed by Juan Benet. It is a P2P distributed file system that "seeks to connect all computing devices with the same system of files to exchange IPFS object" [37]. An object is just a data structure. IPFS can be seen as an amalgam of different internet technologies like Distributed Hash Tables (DHT), the Git versioning system and Bittorrent. All the objects constitute a Merkle Directed Acyclic Graph (DAG) that is a cryptographically verified data structure [30]. In Figure 5, a graph where each node represents an object and all of them constitute a unique file divide on chunks of data is presented. Each object is identified by a hash linked to the starting object.

IPFS is currently being used as a way to relate big files to Ethereum ĐApps. RDF data files can be stored using IPFS and just store a hash pointing to this IPFS resource on the Blockchain.

VII Decentralized Identifiers

According to Zooko Wilcox-O'Hearn triangle diagram, there are three properties that are consider desirable for names participant in a network protocol [38]:

i) **Human-meaningful:** Meaningful and memorable (low-entropy) names are provided to the users.
ii) **Secure:** Any entity in the system can act maliciously, including the majority of the entities or the available computational power.
iii) **Decentralized:** There is still only one, unique and specific entity to which a name resolves.

He conjectured that no single kind of name can achieve more than two properties. For example Bitcoin addresses are secure and decentralized but not human-meaningful. But

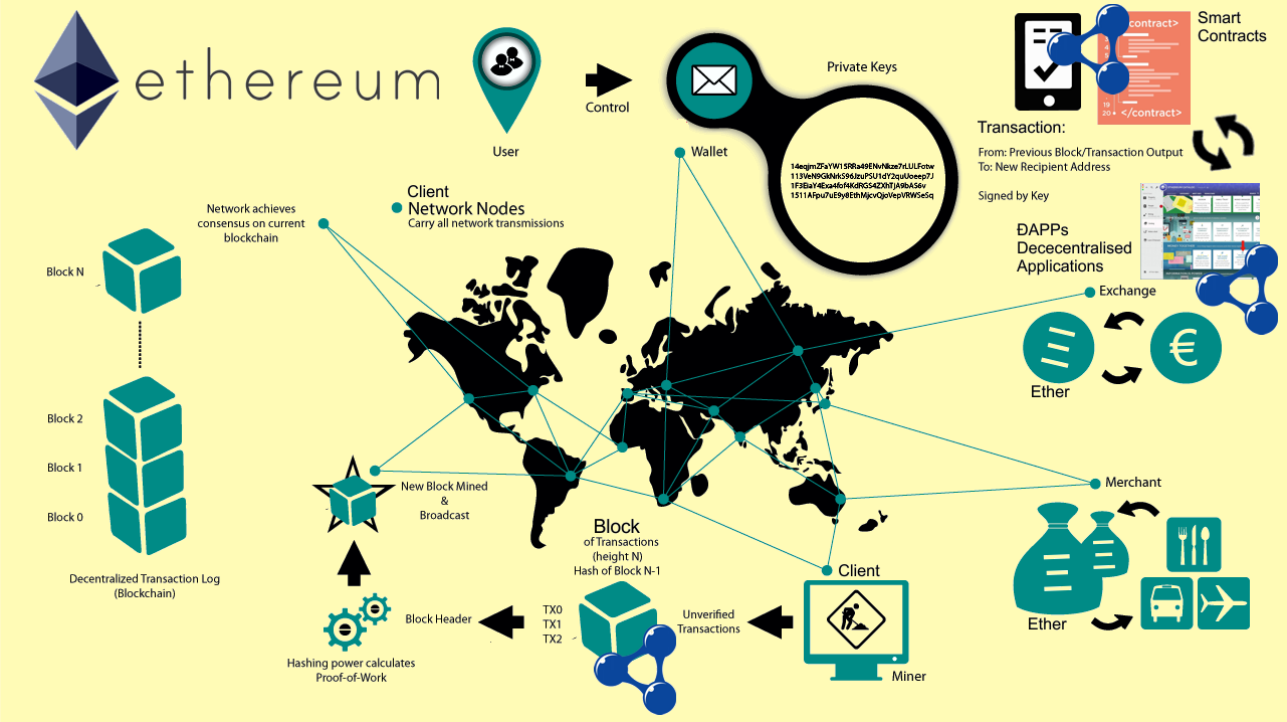


Fig. 6. Ethereum framework elements, modified from [39, p.16]

recently, some systems with the three properties of Zooko's triangle have been created. For instance, Nick Szabo's paper [40] illustrates that all three properties can be achieved.

A Uniform Resource Identifier (URI) is a string of characters used to identify a resource. For example, WebID that is an open standard for identity and login, uses HTTP-URIs. A WebID profile is structured with RDF and uses FOAF vocabulary as a standard. A vulnerable case is when the profile is stored on a personal domain name, and the domain is released Because DNS (Domain Name System) is a centralized component.

One altcoin that gained popularity years ago is NameCoin[15]. It implements a decentralized namespace (.bit domain), therefore a decentralized DNS.

Faisca and Rogado [41] propose an end to end authentication mechanism based on WebID (to represent personal information), JSON Web Tokens (to encode personal related information) and the Blockchain (NameCoin). A "decentralized semantic identity proposal" that uses IPFS to store user profiles.

VIII Semantifying the Blockchain

VIII-A Semantic Blockchain

We have already seen that there is a need for semantic reasoning on the distributed ledger. The natural name for this new addon is "Semantic Blockchain". The Blockchain is the perfect platform to make Semantic web principles widely used and to add to datasets a new property that is "trust". Blockchain logs the truth, or at least the accepted truth by transactions or Smart Contracts interaction, thus these new datasets are completely trustable. Also the Blockchain offers a homogeneous platform to create the new Web 3.0 as opposite of the current heterogeneous web.

This new idea that applies Semantic web technologies on Blockchain-based platforms and/or vice-versa, is currently being under research by few computer scientists and researchers around the world. We proposed three different definitions of Semantic Blockchain:

Definition 6. Semantic Blockchain. *Semantic Blockchain is the use of Semantic web standards on Blockchain-based systems. The standards promote common data formats and exchange protocols on the Blockchain, making used of the Resource Description Framework (RDF).*

Definition 7. Semantic Blockchain. *Semantic Blockchain is a distributed database that maintains a continuously-growing*

[15]https://namecoin.org/

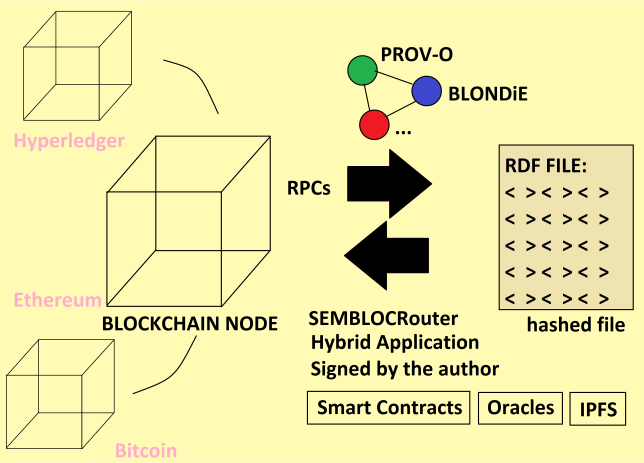


Fig. 7. SemBlocRouter architecture

list of standardized data records, using Resource Description Framework (RDF), hardened against tampering and revision.

Definition 8. Semantic Blockchain. *Semantic Blockchain is the representation of data stored on the distributed ledger using Linked Data.*

VIII-B Semantification process

In fact, there is a big space of research to make in the field of "Semantic Blockchains" or "Semantic Distributed Ledgers", and hopefully it will also impact on the industrial world. For that to happen, it results crucial to start developing applications and frameworks that merge these 2 worlds.

Three possible ways of many to semantify the Blockchain are:

i) Basic mapping of the Blockchain data to RDF making usage of vocabularies, ontologies and remote procedure calls.
ii) Share RDF data directly on the Blockchain, as seen on the section VI, Blockchain data storage is expensive, but maybe saving only hashes pointing to data sets is the solution.
iii) Creation of semantic-ready Blockchains, where all or the core internal data exchange protocols are based on RDF.

SemBlocRouter is a prototype that works as a Hybrid application (Centralized and Decentralized), powered partially by Smart Contracts, where generated RDF data is signed by its owner and file hashes are shared on the network avoiding further data modifications. It uses Oracles for communication with trusted information outside the Blockchain. The files are stored on IPFS and ontologies as BLONDiE and PROV are employed. It is an instant mapper of Blockchain platforms to RDF and offers also a queryable SPARQL endpoint (see Figure 7).

IX Semantic Blockchain Network

The Blockchain is part database, part development platform, part network enabler. It can take many forms of implementations [42]. It is not crazy to think that the world will be

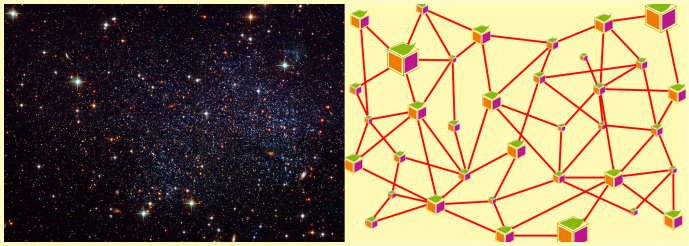


Fig. 8. Universe analogy

composed of thousands running Blockchains, related and interconnected forming a huge network. Each one with thousands or millions of ĐApps (see Figure 8).

This scenario is similar to how the universe is structured where galaxies are composed of systems and each system of planets. A Blockchain is an Artificial Life (AL) agent, life made by man rather than by nature, where rules are defined like it happens on Langton's Ant, Cellular Automata, Von-Neuman self-replicating automaton, Lindenmayer Systems, etc. (see Figure 9).

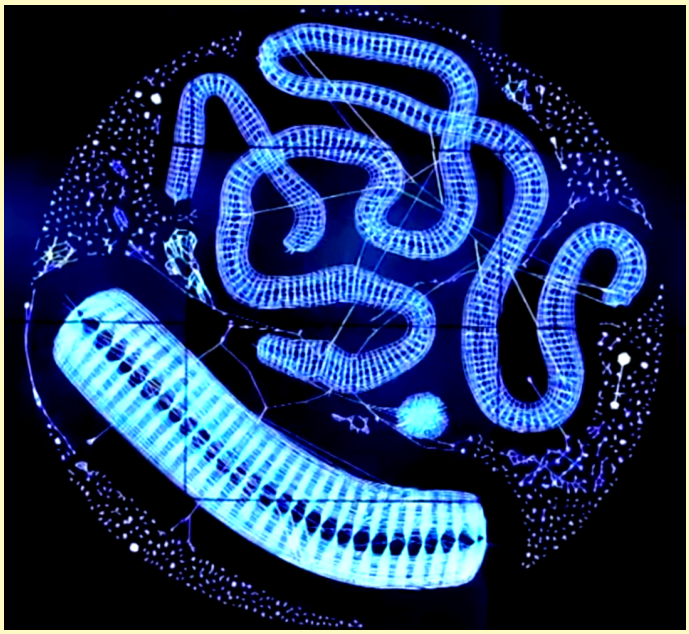


Fig. 9. Historical Bitcoin transactions visualization. Done by: Imperial College Centre for Cryptocurrency Research and Engineering

The usage of Linked Data principles in the different layers of the protocol results crucial in this scenario. A scenario where the following assertions between others should be accomplished:

i) Each Blockchain has to be identified and it should be unique.
ii) Each ĐApp (includes a set of Smart Contacts) is uniquely identified.
iii) Content from a Smart Contract is available also in machine-readable formats (as RDF).

iv) When requested, RDF data from a specific Blockchain is available to other.
v) SPARQL queries compromising different Blockchains should be possible.

X Supply Chain scenario: Decentralized Supply Chain Application (DeSCA)

On 4 June 1982, Keith Oliver coined the term “supply chain” [43]. Stock and Boyer [44] developed a qualitative study to define supply chain management (SCM), summarizing it as “the management of a network of relationships within a firm and between interdependent organizations and business units consisting of material suppliers, purchasing, production facilities, logistics, marketing, and related systems that facilitate the forward and reverse flow of materials, services, finances and information from the original producer to final customer with the benefits of adding value, maximizing profitability through efficiencies, and achieving customer satisfaction” [44].

Linked Data is also being used in Supply Chain Management. There exist some ontologies as The Product Types Ontology[16] and GoodRelations[17]: “The Web Vocabulary for Electronic Commerce” that serve to express data related to e-commerce following the standards that the Semantic Web defines. Also, there is an important work done in the industrial space by the organization GS1 that uses Linked Data standards initiatives on extended packaging, trusted source of data [45].

Current Supply Chain Management systems are heterogeneous. There is a high diversity of solutions that lack interoperability between them. Local ontologies may be considered as enterprise message models [46]. Blockchain per se is a fixer of this reality. If a Blockchain technology or an interconnected network of Blockchains is used for supply chain management, an homogeneous platform will emerge with unique properties.

In a general point of view, Blockchain-based systems can be viewed as token systems. where it is a database with an operation: deduct X units from A and provide X units to B. A has to approved this transaction and has at least X units.

The tokens systems, in general, can be understood as Supply Chain operating systems. Any token is a unit of inventory and an account is an inventory-keeping location, for example a regular store, distribution center or truck trailer. Thus, a supply chain Blockchain application, can be easily used to record balances and transfers of inventory across a distributed supply chain network.

[16]http://www.productontology.org/
[17]http://www.heppnetz.de/projects/goodrelations/

DeSCA is a prototype made to be as most general as possible and simplistic. The main objective is to record the flow of an item across the different participants of a supply chain using the Blockchain architecture where data is also replicated in RDF format.

X-A *Provenance*

W3C Provenance Working Group define provenance as a "record that describes the people, institutions, entities, and activities involved in producing, influencing, or delivering a piece of data or a thing" [47]. The use of the Blockchain as a provenance protocol is a common factor of all use cases, in both scenarios post-trading financial assets or physical entities [48].

Some existing projects working on provenance with Blockchain technologies are available on Table III.

TABLE III
BLOCKCHAIN-BASED PROJECTS WORKING ON PROVENANCE

Project	Scope	Blockchain
Everledger	Diamonds	Eris Stack + Bitcoin
Colu	Digital assets	Bitcoin
Ascribe	Creative digital works	Bitcoin
Monegraph	Creative digital works	Bitcoin
Stampery	Communication, processes and data	Bitcoin
Uproov	Photos, videos and audio recordings	Bitcoin
Provenance	Physical products	Bitcoin, Ethereum

The company project Provenance Ltd[18], based on London and led by Jessi Baker uses Blockchain technology for enabling secure traceability of certifications and other salient data in supply chains. It allows physical products to come with a digital passport proving authenticity and origin by creating auditable records of the journey behind physical products. They take into consideration at each point of time four core properties concerning all materials and consumables: The nature (what it is), the quality (how it is), the quantity (how much of it there is), the ownership (whose it is at any moment) [49].

Everledger[19] founded by Leanne Kemp, is a London start-up focused on diamonds provenance, with a view to expanding into other luxury high-value items whose current provenance relies on paper certificates and receipts. It also provides a Smart Contracts platform to facilitate the transfer of ownership of diamonds to assist insurers in the recovery of items reported as lost or stolen. Each diamond has a unique identifier “a fingerprint”, consisting of 40 different parameters stored on the Blockchain. The 4Cs (color, clarity, cut, and carat weight) that is the universal method for assessing the quality of any diamond is also stored. Modifications of diamonds to change its fingerprints is not a good idea since the cutting produces a lot of wastage. In their first six months, they got more than 850,000 diamonds recorded.

It is relevant to mention that there exists important work related to provenance in the Semantic Web community. Luc

[18]https://www.provenance.org/
[19]http://www.everledger.io/

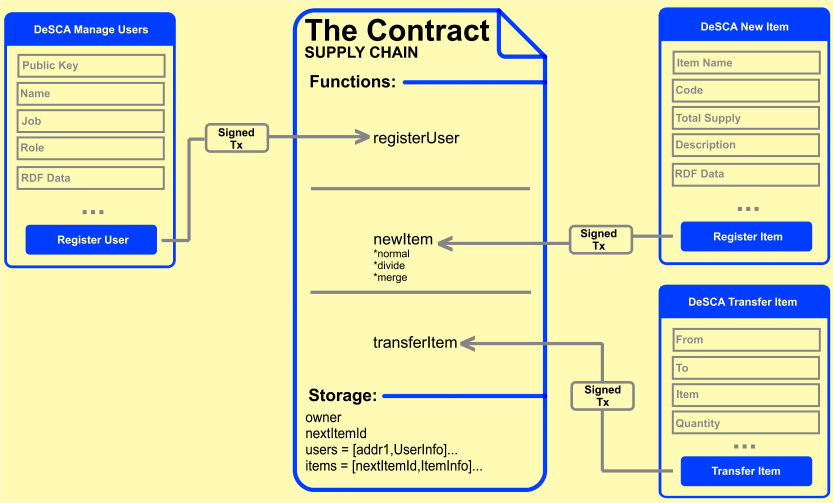


Fig. 10. Smart Contract interaction [36]

Moreau has led different projects related to it such as COLLABMAP and Food Provenance [50]. Moreau et al. also created the ontology PROV-O[20] that serves to represent and interchange provenance information that can be used in different context. Other existing ontologies working with provenance are the lightweight PAV ontology[21]. Christopher Brewster is studying the use of semantic web on agrifood supply chain [51]. He is one of the first proposing the use of Blockchains for this kind of application, in a framework that allows holding semantic data on the Blockchain.

X-B System Architecture

In the software architecture, we want to model the main structures representing the software system and the manner its components communicate among them. The prototype developed is called DeSCA: Decentralized Supply Chain Application is an Ethereum ĐApp. In Figure 11 the native Ethereum's ĐApps architecture is displayed. As explained Ethereum is a P2P framework, the natural division is made on Front-End and Back-End elements. The Back-End is the Blockchain per se, additionally, we consider JavaScript and JQuery logic to handle Front-End elements here. Since we are also using IPFS, its core is also running in the Back-end. A Client running an updated and live node let us interact with the latest state of the Blockchain. The communication between the client and the Front End can be done using existing JavaScript libraries.

The Front End is done using web technologies such as HTML and CSS that work as expected thanks to JavaScript and JQuery code. This code is stored in what we are considering the Back-End. Visual elements are handled by HTML and CSS using the library Bootstrap.

[20]https://www.w3.org/TR/prov-o/
[21]http://pav-ontology.github.io/pav/

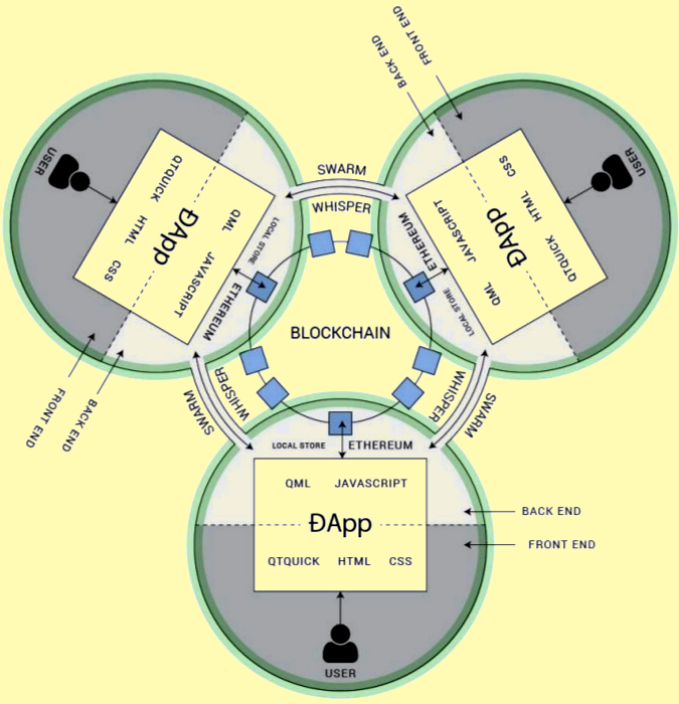


Fig. 11. Ethereum Architecture [52]

X-C Smart Contract Interaction

All the core logic of the decentralized application resides in the Smart Contract coded on Solidity. It is fully possible to interact with the Smart Contract using a client like Geth and specific commands, resulting that the implemented user interface is just a nicer and more user friendly way to interact with it (see Figure 10).

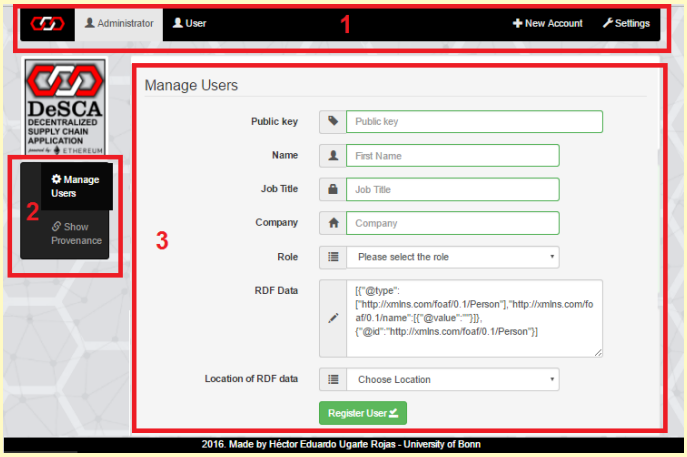


Fig. 12. User interface sections

X-D User Interface and Scenario

The user interface of the prototype is presented in Figure 12. 1) The main menu, where two possible user roles are available (Administrator and Regular User), 2) the local menu according to the selected user, for Administrator: Manage User and Show Provenance, for Regular User: Create Item, Transfer Item and Show Provenance. 3) The respective forms to fill and proceed with the selected task.

A very simple and functional scenario to test DeSCA is the following:

Eric Cartman is a beginning farmer that grows red apples on his farm called "Daisy Hill Puppy Farm" he sells his apples to the company "Umbrella Corp." leaded by Ellie Williams by sacks of 500 apples each. Umbrella Corp. clean, classify, tag and pack the apples by 12 apples per small boxes. They also sell the apple boxes to a local store called "Kwik-E-Mart" owned by Apu Nahasapeemapetilon and they exhibit the apples in their store. Finally Milhouse Van Houten a regular customer of that store buys a box of apples and take it to his home.

All the processes and transactions are registered on the system. The data is also stored in RDF format, where there are 2 possibilities: a) To store data directly on the Blockchain, or b) to store data on IPFS, and share a pointing hash on the contract storage. Finally, once all transactions are registered, the final consumer and any person can verify the registered provenance of the apples.

DEMO: https://youtu.be/JiWDYByK5zo

XI CONCLUSIONS

There are dozens of projects around the globe related to Semantic Web and Blockchain at different levels, some of them were presented on the "W3C Blockchains and the Web Workshop"[22]. Many universities and research centers are focusing on improving and upgrading the Blockchain technology and a few of them are researching on the integration with Linked Data. The main goal of this work is to show that Semantic Web principles can be used on a futuristic Internet that could be composed of Blockchain networks.

First, we started presenting our work done on engineering an ontology, BLONDiE describes inherent Blockchain structure and related information. We also described the EthOn and FIBO ontologies. Later, we stated that linking Blockchains platforms is essential and we described some projects as ILP and Cosmos. Also, we researched about the insertion of semantic content on Smart Contracts, an overview of Ricardian Contracts was given and we concluded that ontologies can help on the design of Smart Contracts. After that, Trust Contracts are introduced as an alternative to be used instead of Turing-complete ones.

The next topic to research is where to store RDF data. Bitcoin and Ethereum allow us to store some data on its Blockchain. These methods are summarized in this paper. Also, we talked about IPFS as an important decentralized storage project. Next, we researched about decentralized identifiers where IPFS hashes, Bitcoin and Ethereum addresses are ideal options here.

The idea of Semantic Blockchain is also discussed and presented, where the process of semantification is possible on current distributed ledgers, SemBlocRouter as an example is introduced. The Semantic Blockchain Network is proposed as a network of Semantic Blockchains with powerful features.

Finally DeSCA (Decentralized Supply Chain Application) is explained where we discuss the importance to have a decentralized supply chain management application in comparison to existing centralized heterogeneous solutions. DeSCA permits the insertion and storage of RDF data that can be read and linked to other data sources.

Acknowledgements

This is just an informal research paper that summarizes some of my work done at the University of Bonn and it was intentionally made to not be part of any journal or something like that. Sorry for the mistakes, typos, etc.

Of course writing it took time, so I am just grateful to my family and all the people mentioned and cited here for researching about amazing and powerful technologies.

I consider that this research paper is just a first step in the integration between these two powerful technologies. Such integration has a colossus potential and there is a huge work to be done that is pretty challenging but at the same time motivating.

Like Nathan Drake said: *"You just count to five and pull the cord. How hard could that be? AHHHHHH! Onetwothreefourfive!"*.

[22] https://www.w3.org/2016/04/blockchain-workshop/

References

[1] N. Shadbolt, T. Berners-Lee, and W. Hall, "The semantic web revisited," *IEEE Intelligent Systems*, vol. 21, no. 3, pp. 96–101, Jan 2006.

[2] bitcoin.org, "Bitcoin developer guide." [Online]. Available: https://bitcoin.org/en/developer-guide

[3] R. Wattenhofer, *The science of the blockchain*. USA: Inverted Forest Publishing, 2016.

[4] M. Swan, *Blockchain : blueprint for a new economy*. Sebastopol, Calif: O'Reilly, 2015.

[5] N. Szabo, "Formalizing and securing relationships on public networks," 1997. [Online]. Available: http://szabo.best.vwh.net/formalize.html

[6] T. Swanson, *Great Chain of Numbers: A Guide to Smart Contracts, Smart Property and Trustless Asset Management*, 2014.

[7] Ethereum, "Create a digital greeter." [Online]. Available: https://www.ethereum.org/greeter

[8] T. Berners-Lee, "Linked data." [Online]. Available: https://www.w3.org/DesignIssues/LinkedData.html

[9] O. Lassila and R. R. Swick, "Resource description framework(rdf) model and syntax specification." [Online]. Available: http://ethdocs.org/en/latest/index.html

[10] A. Hogan, "Linked data and the semantic web standards," in *Linked Data Management (Emerging Directions in Database)*, 1st ed., A. Harth, K. Hose, and R. Schenkel, Eds. Chapman and Hall/CRC, 5 2014, ch. 1, pp. 3–54. [Online]. Available: http://amazon.com/o/ASIN/1466582405/

[11] T. R. Gruber, "A translation approach to portable ontology specifications," *Knowl. Acquis.*, vol. 5, no. 2, pp. 199–220, Jun. 1993. [Online]. Available: http://dx.doi.org/10.1006/knac.1993.1008

[12] N. F. Noy and D. L. Mcguinness, "Ontology development 101: A guide to creating your first ontology," Tech. Rep., 2001.

[13] OWL Working Group, "Owl." [Online]. Available: https://www.w3.org/OWL/

[14] Adam Tinworth, "Next16: Blockchain will build web 3.0, says jamie burke." [Online]. Available: http://nextconf.eu/2016/09/next16-blockchain-will-build-web-3-0-says-jamie-burke/

[15] T. Gerring, "Building the decentralized web 3.0." [Online]. Available: https://blog.ethereum.org/2014/08/18/building-decentralized-web/

[16] S. Tual, "How to get started: your first dapp, under one hour." [Online]. Available: https://forum.ethereum.org/discussion/1402/how-to-get-started-your-first-dapp-under-one-hour

[17] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008. [Online]. Available: http://www.bitcoin.org/bitcoin.pdf

[18] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger homestead draft," 2016. [Online]. Available: http://gavwood.com/Paper.pdf

[19] J. Ziemer, "The blockchain and the elephant. an alignment of ethereum ontology and financial industry business ontology." [Online]. Available: http://finregont.com/2017/02/21/ethereum_fibo_alignment/

[20] A. Abele and J. McCrae, "The linking open data cloud diagram," 2017. [Online]. Available: http://lod-cloud.net/

[21] S. Thomas and E. Schwartz, "A protocol for interledger payments," 2016. [Online]. Available: https://interledger.org/interledger.pdf

[22] J. Kwon and E. Buchman, "Cosmos a network of distributed ledgers," 2017. [Online]. Available: https://github.com/cosmos/cosmos/blob/master/WHITEPAPER.md

[23] G. Wood, "Polkadot: Vision for a heterogeneous multi-chain framework draft 1," 2016. [Online]. Available: https://github.com/polkadot-io/polkadotpaper/raw/master/PolkaDotPaper.pdf

[24] Y. Wand and C. Woo, "Ontology-based rules for object-oriented enterprise modeling," 1999.

[25] H. M. Kim and M. Laskowski, "Towards an ontology-driven blockchain design for supply chain provenance," *CoRR*, vol. abs/1610.02922, 2016. [Online]. Available: http://arxiv.org/abs/1610.02922

[26] I. Grigg, "The ricardian contract," in *Proceedings of the First IEEE International Workshop on Electronic Contracting*, ser. WEC '04. Washington, DC, USA: IEEE Computer Society, 2004, pp. 25–31. [Online]. Available: http://dx.doi.org/10.1109/WEC.2004.19

[27] I. Grigg, "On the intersection of ricardian and smart contracts," 2015. [Online]. Available: http://iang.org/papers/intersection_ricardian_smart.html

[28] OpenBazaar, "Ricardian contracts in openbazaar," 2016. [Online]. Available: https://gist.github.com/drwasho/a5380544c170bdbbbad8

[29] H.-K. Park, C. Park, Y. Choi, and J. H. Choi, "The boscoin white paper," 2017. [Online]. Available: https://steemit.com/cryptocurrency/@boscoin/the-boscoin-white-paper

[30] C. Lundkvist, "Ipfs introduction by example." [Online]. Available: https://www.w3.org/Provider/Style/Data.html

[31] W3C, "Data on the web best practices." [Online]. Available: https://www.w3.org/TR/dwbp/

[32] F. Bauer, *Linked open data: the essentials : a quick start guide for decision makers*. Wien: ed. mono/monochrom, 2012.

[33] T. Berners Lee, "Providing data on the web." [Online]. Available: https://www.w3.org/Provider/Style/Data.html

[34] Ethereum, "Json rpc." [Online]. Available: https://github.com/ethereum/wiki/wiki/JSON-RPC

[35] Colored-Coins, "Data storage methods - data storage on the blockchain," 2016.

[36] H. Ugarte, "Strategies for integrating semantic and blockchain technologies," 2016. [Online]. Available: https://es.slideshare.net/hedugaro/strategies-for-integrating-semantic-and-blockchain-technologies

[37] J. Benet, "Ipfs - content addressed, versioned, p2p file system (draft 3)," 2016. [Online]. Available: https://ipfs.io/ipfs/QmR7GSQM93Cx5eAg6a6yRzNde1FQv7uL6X1o4k7zrJa3LX/ipfs.draft3.pdf

[38] Z. Wilcox-O'Hearn, "Names: Distributed, secure, human-readable: Choose two." [Online]. Available: https://web.archive.org/web/20011020191610/http://zooko.com/distnames.html

[39] A. Antonopoulos, *Mastering bitcoin : unlocking digital cryptocurrencies*. Sebastopol, CA: O'Reilly, 2015.

[40] N. Szabo, "Secure property titles with owner authority." [Online]. Available: http://nakamotoinstitute.org/secure-property-titles/#selection-11.0-11.10

[41] J. G. Faísca and J. Q. Rogado, "Decentralized semantic identity," in *Proceedings of the 12th International Conference on Semantic Systems*, ser. SEMANTiCS 2016. New York, NY, USA: ACM, 2016, pp. 177–180. [Online]. Available: http://doi.acm.org/10.1145/2993318.2993348

[42] W. Mougayar, *The Business Blockchain: Promise, Practice, and Application of the Next Internet Technology*, 1st ed. Wiley, 5 2016. [Online]. Available: http://amazon.com/o/ASIN/1119300312/

[43] P. Heckmann, D. Shorten, and H. Engel, "Supply chain management at 21 the hard road to adulthood."

[44] J. R. Stock and S. L. Boyer, "Developing a consensus definition of supply chain management: a qualitative study," *International Journal of Physical Distribution & Logistics Management*, vol. 39, no. 8, pp. 690–711, 2009. [Online]. Available: http://dx.doi.org/10.1108/09600030910996323

[45] M. Harrison, "Semantic web technologies introduction and relevance to gs1 community," 2015.

[46] M. Zdravković, M. Trajanović, and H. Panetto, "Local ontologies for semantic interoperability in supply chain networks," 2011.

[47] L. Moreau and P. Missier, "Prov-dm: The prov data model." [Online]. Available: https://www.w3.org/TR/2013/REC-prov-dm-20130430/#dfn-provenance

[48] I. Allison, "Provenance has a big year ahead delivering supply chain transparency with bitcoin and ethereum." [Online]. Available: http://www.ibtimes.co.uk/provenance-has-big-year-ahead-delivering-supply-chain-transparency-bitcoin-ethereum-1537237

[49] J. Steiner and J. Baker, "Blockchain: the solution for transparency in product supply chains." [Online]. Available: https://www.provenance.org/whitepaper

[50] L. Moreau, "Enabling provenance on the web. standardization and research questions." [Online]. Available: http://internet-conf.org/wp-content/uploads/2015/10/lucmoreau-provenance-2015.pdf

[51] C. Brewster, "Semantic blockchains in the supply chain." [Online]. Available: http://www.slideshare.net/christopherbrewster/20150617-tno

[52] G. Wood, "The ethereum experience." [Online]. Available: http://www.slideshare.net/ethereum/the-ethereum-experience