

# Semantic Blockchain Technical Report

Firas Kassawat, Ernane Luis, S. Matthew English  
Supervisor:  
Prof. (Univ. Simón Bolívar) Dr. Maria Esther Vidal

MA-INF 4314

# Contents

0.1	Introduction . . . . .	2
0.2	Bitcoin & Blockchain . . . . .	2
0.2.1	What is “Semantic Blockchain”?	3
0.3	Scenario Presentation & Motivation . . . . .	4
0.4	Timeline . . . . .	5
0.5	Requirements & Testing Framework . . . . .	6
0.5.1	How To Run Toshi . . . . .	8
0.6	System Architecture . . . . .	9
0.6.1	Unification of Different Resources . . . . .	9
0.6.2	Why Graph Database? . . . . .	10
0.7	Integration System . . . . .	12
0.8	Semantic Blockchain . . . . .	12
0.8.1	Blockchain Ontology Survey . . . . .	13
0.9	User Case . . . . .	13
0.9.1	Knowledge Graph . . . . .	13
0.9.2	Data Analytics . . . . .	14
0.9.3	Query Processing . . . . .	14
0.10	Future Work . . . . .	14
0.11	Concluding Discussion . . . . .	15

## 0.1 Introduction

The system for online value transmission known as Bitcoin (BTC) initiated research into the practical applicability of so-called blockchain technology as a mechanism for the creation of what has been dubbed a “permissionless and censorship-resistant” protocol for information interchange involving transactions of financial and social capital. The security and privacy characteristics of every technology develop dynamically based on the demands of its users. Ultimately it’s up to each individual actor to determine their own optimal balance between the need for confidentiality and the need for convenience. A common assumption is that Bitcoin is anonymous. This is incorrect as transactions can be de-anonymized, limiting the commercial utility of the network and also undermining individual privacy, however in some cases user’s may value a voluntary association with a particular Bitcoin address. In this situation a user trades privacy for convenience, not an unusual bargain in the development of the modern web. In this work we utilize information that associates the Twitter handle of an individual user with a particular Bitcoin address towards the creation of a federated knowledge graph.

In its capacity as a medium of value exchange Bitcoin presents a new model for research into the causes of economic distortions that create risk. The novelty of the system is predicated on the fact that the complete set of transactions in the Bitcoin ecosystem is recorded by the blockchain, a publicly available information resource [8]. All BTC exists as divisible and compoundable units assigned to specific addresses indexed in the global list of Unspent Transaction Output (UTXO).

## 0.2 Bitcoin & Blockchain

Bitcoin is a complex protocol. We provide here a brief sketch of the components necessary to understand the analysis presented in this chapter, however due to the many moving parts of the system this is necessarily a superficial overview. Interested parties are referred to [11] for a more complete picture of the system.

The decentralized currency protocol known as Bitcoin was proposed by Satoshi Nakamoto [10]. The system utilizes a peer-to-peer (P2P) architecture that enables users to send and receive transactions denominated in units of BTC. Users are represented in the network by a public/private key pair. Units of BTC can be transmitted to a user by specifying a hash of that user’s public key as the receiving party, providing a degree of pseudo-anonymity. Users can generate many public keys, i.e. receiving addresses. The corresponding private keys are used to sign (authorize) transactions. Private keys are stored in a “wallet” either locally or provided as a hosted service.

To participate in the Bitcoin network the user runs a client software, such as the Satoshi client, which communicates with a set of peers. Transactions are broadcast by the Bitcoin client and received by the peer-to-peer network. They are confirmed after having been added to the “blockchain” - similar to a linked

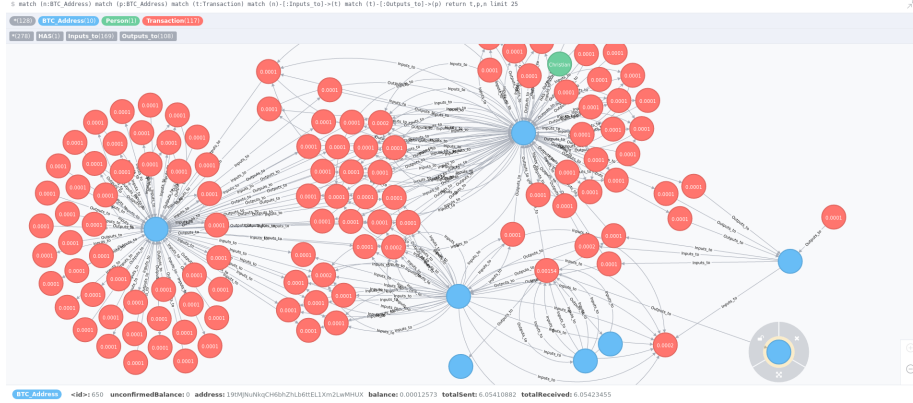


Figure 1: Knowledge Graph Structure

list with the subtle difference that it references the previous block using its hash rather than a pointer. This data structure contains blocks of all accepted transactions since the genesis of the system [17].

Every full node running a Bitcoin client maintains a complete copy of this public blockchain. The block generation process confirms new transactions. It necessitates the satisfaction of a computationally expensive “proof of work” puzzle. A valid solution to this puzzle entitles the party that delivers it to the wider network the privilege to issue themselves a reward in the form of newly minted coins. The information available through the graph structure of the Bitcoin P2P network is limited due to the dynamic block formation process. Each node only has direct knowledge of the peers to which its client is connected. The graph of all transactions can be constructed entirely from the publicly available blockchain, wherein the nodes of the graph correspond to Bitcoin addresses and the edges to transactions performed between those addresses.

### 0.2.1 What is “Semantic Blockchain”?

Semantic Blockchain is an emerging paradigm in database design and development. It describes a model of information repository that incorporates the distributed consensus mechanism popularized by cryptocurrency implementations, such as Bitcoin, and the exchange protocols of the linked open data specification. Applications are constructed to support semantic queries and elements of logical reasoning [5]. Semantic blockchain principles are integral to the web of interlinked blockchains and the associated features, such as decentralized exchange.

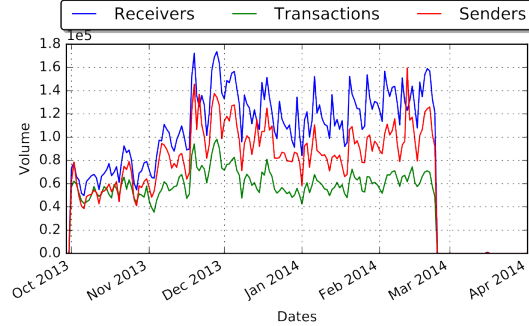


Figure 2: Network Analytics

### 0.3 Scenario Presentation & Motivation

Bitcoin is a digital asset and value transmission system. The Bitcoin ecosystem is described by its openness, its distributed (peer-to-peer) organizational model, and the fact that by default addresses, i.e. users, are pseudo-anonymous [12]. Sophisticated network analytics through data integration have heretofore been infeasible due to the limitations of available blockchain explorers, such as `blockchain.info` & `blockexplorer.com`. One of the primary problems encountered by users of blockchain explorers when conducting analyses of the Bitcoin transaction network are its resistance to compliance with regulatory requirements such as anti-money laundering (AML) & know your customer (KYC) [1]. The relative inefficiencies of existing solutions can be described as follows:

- No issuing of queries
- No clustering of addresses
- No association of pseudonyms with “reality”
- No voluntary association with online identity

Regulators are interested in finding a way to associate Bitcoin addresses with real-world identities. Simultaneously a large subset of user’s has already voluntarily associated their own Bitcoin address with a Twitter handle in such a way that the validity of this assertion is established by posting a unique code to through the Twitter handle of the account in question [6]. This dataset is known as OneName<sup>1</sup>. It is described as “the easiest way to start receiving bitcoins”. Indeed it clearly establishes a channel whereby an address, i.e. “Bitcoin username”, can be related to a person and thereby enable people to send them Bitcoins. When involving transactions that require trust—such as those across the Bitcoin transaction network it is beneficial to have an identity and a reputation associated with a particular address. Onename is a way for people to

<sup>1</sup><http://onename.com>

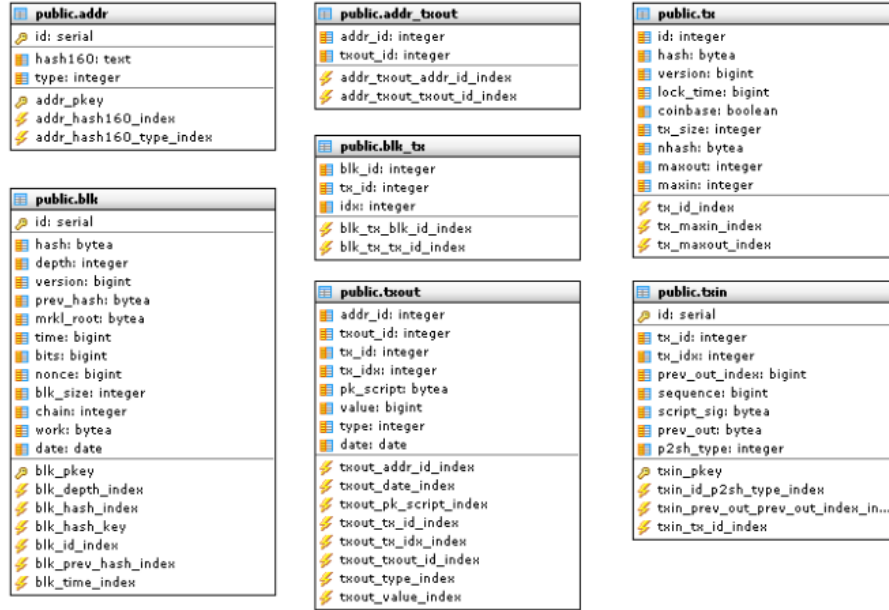


Figure 3: Relational Database

voluntary associate an ID with a bitcoin address, validated by their Twitter account [2].

## 0.4 Timeline

This project was conceived of and proposed in the Spring of 2016. Over the course of it's duration the following tasks were executed, chronologically. The work commenced with a meeting together with the project mentor in office and all group participants. This resulted in the preparation of a use case. Initial results were documented are prepared for presentation via slide show. A follow-up meeting with the group mentor in office was planned. This precipitated a programme of research about relevant tools and architecture principles. Subsequently the next meeting with a tutor was organized following an internal discussion of "brain storming session". The net result of this was documented and prepared for presentation, detailing the system requirements.

Internal meeting and setting up server for extract data from the blockchain. Server up and running and data collected, first implementation for data wrappers and data collectors. Second sprint of implementation of the data API wrappers and data collectors from twitter and blockchain data. Visualization of transaction from one node to another using D3 library. Sketech of the main code and wrappers. Data Server Evaluation and SPARQL Testing. Validating

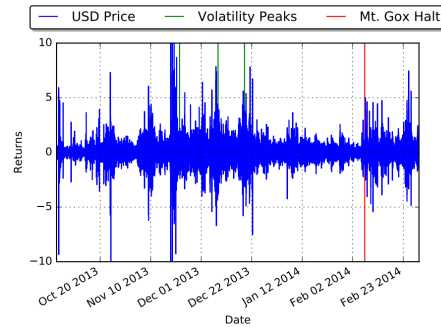


Figure 4: Bitcoin Price Fluctuations

wrappers with Toshi Database. Found out its impossible to get the senders Data from IT. Testing Abe To use instead of Toshi. Result: It is possible to get senders data from Abe. Modeling a new knowledge Graph to represent the new linked data and solving our issues with current data. Web Development to acquire the work we have done and embed it in our website. Create a Cypher file for populating the Knowledge Graph. Website finishing touches. Test population and website applications. Adding Linked data with Twitter. Performing statistics queries. Documenting our result so far.

Project Documentation and testing website template design. Django Website deployment. Graph database transactional data and network analysis. On the fly data integration in implementing admin privileges for data visualization. Transactional data using graph db, Django app. Adapting BLONDiE extension. Formatting ontology extension. RDF OWL Schema for the knowledge graph. Web page for the URIs resources and using RDF HTML and/OR basic RDF. Publish the Website with domain and update Data.

## 0.5 Requirements & Testing Framework

Import Data from the Blockchain and include URIs and relate to the OWL schema. Visualization Bostok Chart. Add the register and login with BTC address. Add a form to get transactions for a certain period of time return details of this data and graphs for it. Web page for the URIs resources and using RDF HTML and/OR basic RDF. Publish the Website with domain and update Data. Documentation of the Poster and the Presentation. Provide Endpoint Cypher and REST. Adding template Queries for cypher and any other endpoint. RDF OWL Schema for the knowledge graph. Set Up the D2RQ server at our Server. Links and Server Commands.

### SERVER DETAILS:

```
ip: 46.101.180.63
user: admin
password: uni-bonn
```

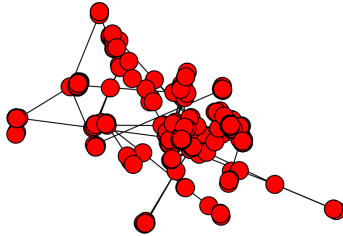


Figure 5: Maximally Connected Graph Component

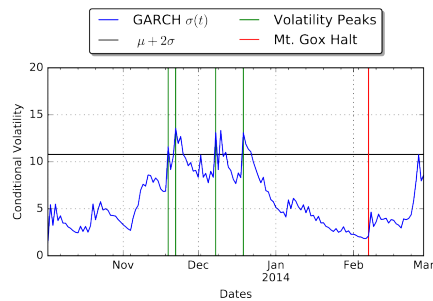


Figure 6: Behaviour Snapshot of Bitcoin Transaction Graph

To connect run this command, all the files are inside the Folder **sabo ssh**  
**admin@46.101.180.63** There are 3 Dockers running for Toshi

1 - Postgres docker  
 IP: 172.17.0.2:5432  
 user: postgres  
 password: <no-password>

2 - Regis docker:6379  
 IP: 172.17.0.3

After Toshi have download data from the blockchain we can use this command  
 at the server to download the **postgres.sql** data/

```
docker exec -t toshi_db pg_dump -U
postgres -h 172.17.0.2 -p 5432 postgres > postgres1.sql
```

After running this command at you local machine to download from the remote  
 server:

```
scp admin@46.101.180.63:sabo/postgres1.sql .
```



## Notable Addresses

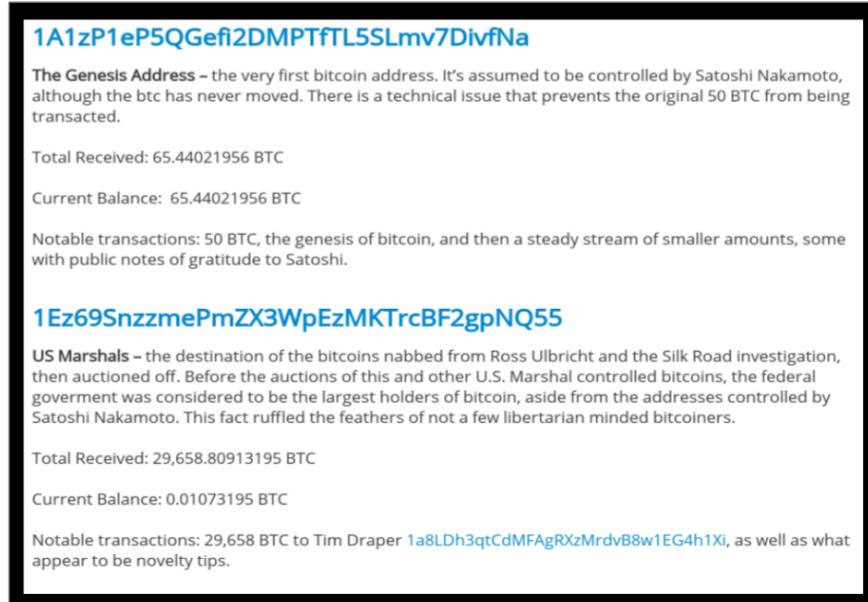


Figure 7: Accounts that Warrant Close Scrutiny

### 0.5.1 How To Run Toshi

This command is to use for the first time to create the database schema:

```
docker run --name toshi_main -d -p 5000:5000 -e
REDIS_URL=redis://172.17.0.3:6379 -e
DATABASE_URL=postgres://postgres:@172.17.0.2:5432 -e
TOSHI_ENV=development coinbase/toshi sh -c 'bundle exec
rake db:create db:migrate; foreman start'
```

After this command at other times you can use this command to start the Toshi.

```
docker run --name toshi_main -d -p 5000:5000 -e
REDIS_URL=redis://172.17.0.3:6379 -e
DATABASE_URL=postgres://postgres:@172.17.0.2:5432 -e
TOSHI_ENV=development coinbase/toshi sh -c 'foreman start'
```

Create the Mapping File for D2RQ. Creating a Docker container. Create some useful case using our SPARQL Endpoint. Visualization and subgraph mining. One server up and running with Abe to get the blockchain into a relational database. Collect 200MB of Testnet3 data as RD from PostgreSQL. Creating “pop-up” Box on Visualization Name. Find a service that connects RDF to an endpoint. We have used D2RQ <http://d2rq.org/> for mapping our Relational Database to SPARQL. Download the original Bitcoin Blockchain (at

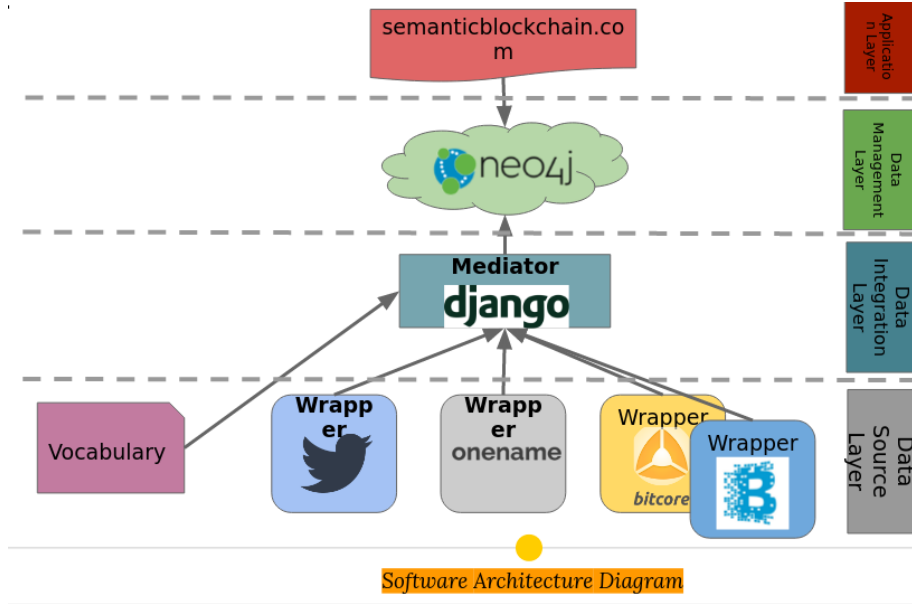


Figure 8: Architecture of System Components

least 1GB) Data. One programme that is passing the relational database to export it as RDF. Setup Domain at server. Setup Server.

## 0.6 System Architecture

### 0.6.1 Unification of Different Resources

Queries that can aggregate these disparate information resources (Twitter, One-Name, & the Bitcoin transaction network) would provide seamless integration of data. This in turn facilitates the creation of a more complete picture of network participants and interactions [18]. The question then arises of how to store this aggregated information? The relational database model is not well suited to the creation of a knowledge graph that draws from the information resources we herein consider. One of the main deficiencies of the relational database model is that the sender address is not specified, due to the UTXO structure of a Bitcoin transaction.

Deficiencies of Relational Database Model

```
SELECT tx_id , tx_version , tx_lockTime , tx_size
FROM tx
WHERE tx_hash =
          9bb5071f1b1c3ef3bb3fcf2497712c86178bdb0bd047dc4103233979a55ec2e7
join
```

```

SELECT
txout.txout_pos ,
txout.txout_scriptPubKey ,
txout.txout_value ,
nexttx.tx_hash ,
nexttx.tx_id ,
txin.txin_pos ,
pubkey.pubkey_hash
FROM txout
LEFT JOIN txin ON (txin.txout_id = txout.txout_id)
LEFT JOIN pubkey ON (pubkey.pubkey_id = txout.pubkey_id)
LEFT JOIN tx nexttx ON (txin.tx_id = nexttx.tx_id)
WHERE txout.tx_id = ?
ORDER BY txout.txout_pos
join
SELECT
txin.txin_pos ,
txin.txin_scriptSig ,
txout.txout_value ,
COALESCE(prevtx.tx_hash , u.txout_tx_hash) ,
prevtx.tx_id ,
COALESCE(txout.txout_pos , u.txout_pos) ,
pubkey.pubkey_hash
FROM txin
LEFT JOIN txout ON (txout.txout_id = txin.txout_id)
LEFT JOIN pubkey ON (pubkey.pubkey_id = txout.pubkey_id)
LEFT JOIN tx prevtx ON (txout.tx_id = prevtx.tx_id)
LEFT JOIN unlinked_txin u ON (u.txin_id = txin.txin_id)
WHERE txin.tx_id = ?
ORDER BY txin.txin_pos

```

Can be reduced to the following:

```

Match (b:BTC_Address) -[:recieve]
->(p:Transaction {hsh:{hsh}}) -[:recieve]
->(b:BTC_Address)

```

## 0.6.2 Why Graph Database?

The graph database model can be considered as a natural representation of highly inter-connected data, such as the network of Bitcoin transactions or Twitter users [7]. Special graph storage structure. Efficient schema-less graph algorithms. Support for query languages. Operators to query the graph structure. Best way to represent transactional data similar to Blockchain data [9].

Regarding *Triple Store*. The goal is to facilitate query processing & data analytics. Global identifier to establish the integration. Regarding *Transactional*

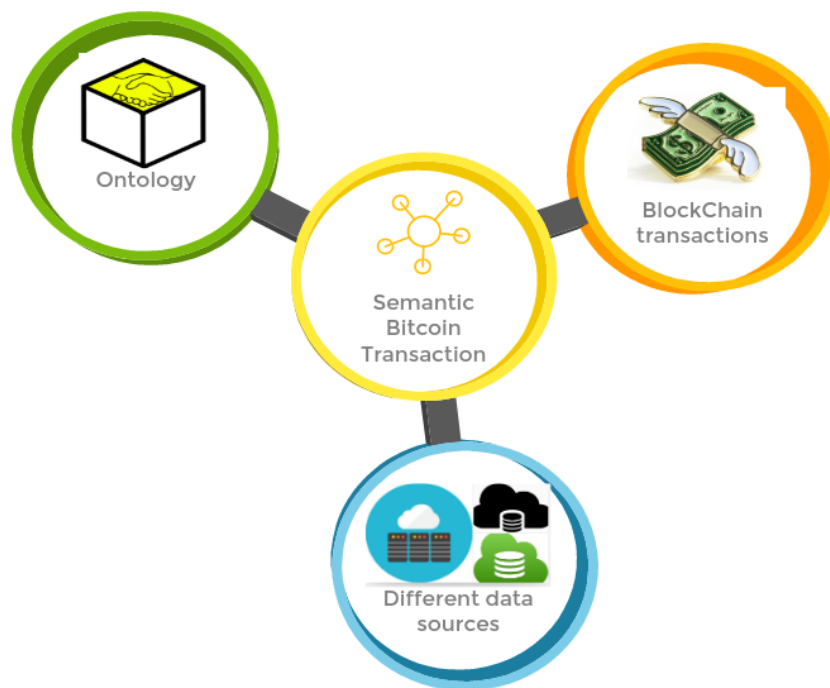


Figure 9: High Level Design Structure

*Graph Database* Representing Transactions Graph Data For traversal look up and querying is faster:

- Human readable representation
- Issue queries of high complexity
- Efficient query processing throughout graph
- Low disk space

## 0.7 Integration System

Integration System

$IS = \langle O, S, M \rangle$

O: Global Schema Defined by the  
Vocabulary (Knowledge Graph)  
S: List of Acquired Data Sources  
M: List Mapping the Data from  
Sources to Global Schema (ETLs)

## 0.8 Semantic Blockchain

### Why Neo4J?

Intuitive Using a graph model for data representation [3]. Fast With a powerful traversal framework for high-speed graph queries Massively scalable Native graph storage Up to several billion nodes/relationships/properties Simple Accessible by a convenient REST and Web interface Expressive With a powerful, human readable graph query language (Cypher) Graph Analysis Plugins Graph Compute and many other plugins that provide additional assistance in graph analysis.

O: Global schema Vocabulary

RDF RDFS Schema.org OWL  
Blondie Sabo

RDF, RDFS, OWL

What did we include ?

rdf:type  
rdf:property

owl:Class

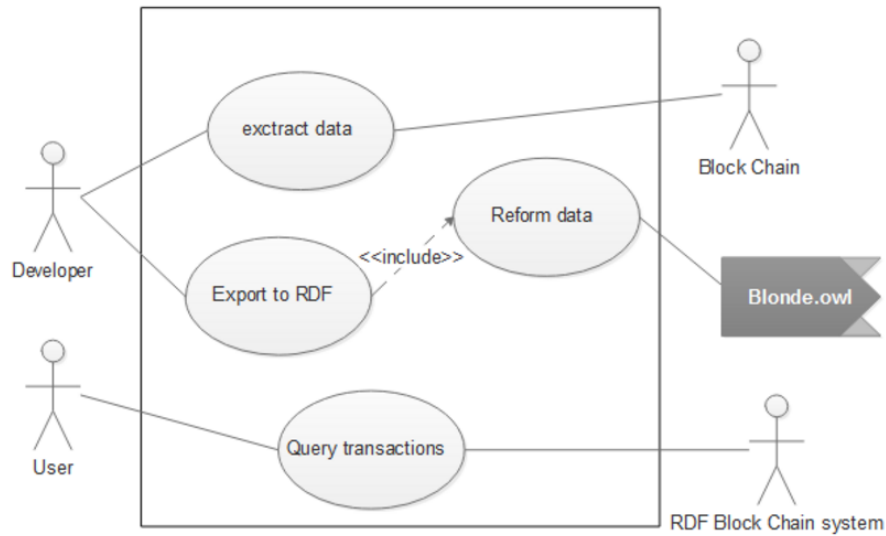


Figure 10: Use Case Diagram

owl:ObjectProperty  
..

rdfs:subclassof rdfs:range rdfs:domain

### 0.8.1 Blockchain Ontology Survey

Existing vocabularies capture basic transaction structure, such as [melvincarvalho/crypto-currency-ontology](#) [4] don't generalize beyond basic Bitcoin architecture. Blockchain Ontology with Dynamic Extensibility (BLONDiE). Incorporates alt coins, particularly Ethereum [15]. Extended integrate with diverse range of sources.

## 0.9 User Case

We Image 3 possibles user cases for ou tool: Knowledge Graph, Query Processing and Data Analytics.

### 0.9.1 Knowledge Graph

We would like to analyze the Bitcoin blockchain data in a graphical way, so for that to be possible to answers questions like: Are a twitter user sending bitcoins to any of their follwers? or What are the transactions between two users? What

**Home** Welcome to Blockchain More...

Height	Age	Transactions	Total Sent	Relayed By	Size (kB)
410168	5 minutes	2545	24,772.91 BTC	BitFury	998.08
410167	22 minutes	2090	17,225.59 BTC	AntPool	998.05
410166	32 minutes	3098	32,947.95 BTC	AntPool	998.17
410165	56 minutes	2024	22,037.33 BTC	AntPool	998.07
410164	1 hour 6 minutes	2691	34,444.14 BTC	AntPool	998.11
410163	1 hour 28 minutes	2351	20,429.93 BTC	BitFury	998.19

**Latest Transactions**

b82a41ae50c51ee8e405c50cf...	< 1 minute	0.07534672 BTC
2d01b63383eb033d92d8d2d...	< 1 minute	0.01348473 BTC
b30a77e77d5145fa10509f74...	< 1 minute	3.8337595 BTC
4a927d338a338ad4a4a9b70f...	< 1 minute	2.46 BTC
882d2149e6dc3a68d84cf4a5c...	< 1 minute	0.49459964 BTC
88d8c5221d9554b4d6da28c5...	< 1 minute	12.96774727 BTC
02599d8e577d978051ec69a0...	< 1 minute	72.39101795 BTC

**Search**  
You may enter a block height, address, block hash, transaction hash, hash160, or ipu4 address...

Address / ip / SHA hash

**NEWS**

Magnr - Bitcoin Trading Platform | Trade with Leverage  
Magnr - 1 minute ago

€500, the 'Bin Laden' of banknotes, could be axed  
(Blockchain 4 minutes ago)

Gavin breaks Twitter silence: 'in my head today: Truth is stranger than fiction. And: People are weird.'  
(Bitcoin 27 minutes ago)

Figure 11: Typical Block Explorer [14]

is the shortest path between two user over the transaction cloud? or How users of the Bitcoin are related?

## 0.9.2 Data Analytics

A second possible user case is analyze the raw data from the blockchain with the purpose of drawing conclusions about that information. Data analytics is used in many industries to allow companies and organization to make better business decisions and in the sciences to verify or disprove existing models or theories [16].

## 0.9.3 Query Processing

Using the power of Graph database we could traverse the data and do better Human readable queries over the blockchain raw data. Powering by Cypher Query Language from Neo4J and comparing againsts SQL Query from Toshi PostgreSQL Schema we could get optimal results of queries:

## 0.10 Future Work

We define our future work by: adding in our graph database all the bitcoin blockchain data Support for more comprehensive semantics [13] make the blockchain data available in the Linked Data Cloud Add more public data into our Knowledge graph to get more insights

## 0.11 Concluding Discussion

With a total market capitalization in excess of \$10,000,000,000, more than 7 years of continued operation, and over 1.9 million addresses Bitcoin is the world's most successful blockchain-based cryptocurrency. We argue that it is the world's *only* long-term viable blockchain application to date. This work comprises a comprehensive programme of research undertaken into the nature of the Bitcoin protocol in an effort to assess the degree to which it's fundamental components could be applied to alternative use-cases. The question we sought to answer was whether the data structures and incentive mechanisms that facilitate Bitcoin could be melded with semantic applications and techniques to realize concrete solutions to practical problems. Having demonstrated this result in the affirmative we look forward to the continued growth of this burgeoning field of research.



# Bibliography

- [1] Kyle Croman, Christian Decker, Ittay Eyal, Adem Efe Gencer, Ari Juels, Ahmed Kosba, Andrew Miller, Prateek Saxena, Elaine Shi, and Emin Gün. On scaling decentralized blockchains. In *Proc. 3rd Workshop on Bitcoin and Blockchain Research*, 2016.
- [2] Christian Decker and Roger Wattenhofer. Information propagation in the bitcoin network. In *IEEE P2P 2013 Proceedings*, pages 1–10. IEEE, 2013.
- [3] Juan Garay, Aggelos Kiayias, and Nikos Leonardos. The bitcoin backbone protocol: Analysis and applications. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 281–310. Springer, 2015.
- [4] David Garcia, Claudio J Tessone, Pavlin Mavrodiev, and Nicolas Perony. The digital traces of bubbles: feedback cycles between socio-economic signals in the bitcoin economy. *Journal of the Royal Society Interface*, 11(99):20140623, 2014.
- [5] Stuart Haber and W Scott Stornetta. How to time-stamp a digital document. In *Conference on the Theory and Application of Cryptography*, pages 437–455. Springer, 1990.
- [6] Adrian Hope-Bailie and Stefan Thomas. Interledger: Creating a standard for payments. In *Proceedings of the 25th International Conference Companion on World Wide Web*, pages 281–282. International World Wide Web Conferences Steering Committee, 2016.
- [7] Sepandar D Kamvar, Mario T Schlosser, and Hector Garcia-Molina. The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th international conference on World Wide Web*, pages 640–651. ACM, 2003.
- [8] ROBERT McMillan. The inside story of mt. gox, bitcoin’s \$460 million disaster. <http://www.wired.com/2014/03/bitcoin-exchange>, 2014.
- [9] Malte Möser, Rainer Böhme, and Dominic Breuker. Towards risk scoring of bitcoin transactions. In *International Conference on Financial Cryptography and Data Security*, pages 16–32. Springer, 2014.

- [10] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2008.
- [11] Arvind Narayanan, Joseph Bonneau, Edward Felten, Andrew Miller, and Steven Goldfeder. *Bitcoin and cryptocurrency technologies*. Princeton University Press, 2016.
- [12] Micha Ober, Stefan Katzenbeisser, and Kay Hamacher. Structure and anonymity of the bitcoin transaction graph. *Future internet*, 5(2):237–250, 2013.
- [13] Frederick Reese. Rethinking bitcoin’s \$10 billion market cap. *CoinDesk*, July 3, 2016.
- [14] Fergal Reid and Martin Harrigan. An analysis of anonymity in the bitcoin system. In *Security and privacy in social networks*, pages 197–223. Springer, 2013.
- [15] Dorit Ron and Adi Shamir. Quantitative analysis of the full bitcoin transaction graph. In *International Conference on Financial Cryptography and Data Security*, pages 6–24. Springer, 2013.
- [16] Evan Schwartz. A payment protocol of the web, for the web: Or, finally enabling web micropayments with the interledger protocol. In *Proceedings of the 25th International Conference Companion on World Wide Web*, pages 279–280. International World Wide Web Conferences Steering Committee, 2016.
- [17] Nick Szabo. The idea of smart contracts. *Nick Szabos Papers and Concise Tutorials*, 1997.
- [18] Stefan Thomas and Evan Schwartz. A protocol for interledger payments. URL <https://interledger.org/interledger.pdf>, 2015.