Datawallet - A blockchain-based wallet for facilitating exchanging, tracking and tracing of data

Kazé A. ONGUENE Rishav Anand

BRAINCITIES LAB
21 Boulevard Haussmann, 75009 Paris, France

Contents

Disclaimer

This White Paper is for information purposes only. BRAINCITIES LAB does not make and expressly disclaims all representations and warranties, express, implied, statutory or otherwise, whatsoever, including, but not limited to: (i) warranties of merchantability, fitness for a particular purpose, suitability, usage, title or non-infringement; (ii) that the contents of this white paper are free from error; and (iii) that such contents will not infringe third-party rights. BRAINCITIES LAB and its affiliates shall have no liability for damages of any kind arising out of the use, reference to, or reliance on this white paper or any of the content contained herein, even if advised of the possibility of such damages. BRAINCITIES LAB does not guarantee the accuracy of or the conclusions reached in this white paper, and this white paper is provided “as is”. In no event will BRAINCITIES LAB or its affiliates be liable to any person or entity for any damages, losses, liabilities, costs or expenses of any kind, whether direct or indirect, consequential, compensatory, incidental, actual, exemplary, punitive or special for the use of, a reference to, or reliance on this white paper or any of the content contained herein, including, without limitation, any loss of business, revenues, profits, data, use, goodwill or other intangible losses. This paper is a description of the current and planned BRAINCITIES LAB ecosystem, the actors building it and the project that is working on it. It is neither a solicitation nor a prospectus. This paper may include predictions, estimates or other information that might be considered forward-looking. While these forward-looking statements represent BRAINCITIES LAB’s current judgment on what the future holds, they are subject to risks and uncertainties that could cause actual results to differ materially. You are cautioned not to place undue reliance on these forward-looking statements, which reflect the opinions of BRAINCITIES LAB only as of the date of this publication. Please keep in mind that BRAINCITIES LAB is not obligating itself to revise or publicly release the results of any revision to these forward-looking statements in light of new information or future events.

Abstract

Since the boom of the internet, users have been generating data more than ever and service providers have been using this abundance of data to create data-based products and to deliver tailored content back to the user. There exist two main entities in this data exchange, the data owner and the service provider and they both have their problems. On the data owner's side, their data, which contains a lot of their personal information, is being used and being sold without their consent or without getting any profits from the sale. On the service provider's side, there is a lack of latest, accurate and verifiable data which is the key to proactive decision making. There is a clear need for a data storage and exchange system where owners can store and exchange data with the consumers in a consensual manner. This system also needs to provide data tracking to allow users to see which piece of their information is shared with which service provider. For service providers, it also needs to provide data traceability to facilitate the supply of latest, accurate and verifiable data in return for incentives. In this whitepaper, we work out "Datawallet" - a blockchain-based wallet for facilitating exchanging, tracking and tracing of data. We also list down various use cases that focus on tracing, tracking and incentivising data owners during a data exchange using Datachain Tokens(DCT).

1. Introduction

With the creation of more and more data-based products, the value of data is rising. It is time that data owners look at their data in the same way as they look at their money. Data needs to be tracked and shared with the same precaution that is taken with fiat money. This generates a clear need for a *data bank* capable of providing a secure way of sending, receiving and tracking data for the data owners and also a *data market* capable of facilitating the exchange of accurate and verifiable data in bulk between data owners and service providers for some incentives.

Similar to traditional banks, a data bank is a virtual bank for your data. Data banks should allow their members to audit every piece of data they own. Then, it should provide them with the tools for sending and receiving data directly to and from their data accounts. They should have complete control over which piece of data they are allowed to share instead of having to share the whole data document/file. Further, it should allow users to track their data to audit which user/service holds what and how much of their data.

The need for a data exchange comes into the picture for service providers when they want to access bulk, accurate and verified data. This data may be used for data-based products or for providing greater user experience back to the data owner. The problem with data-based products is that you do not see your data inside it in a regular text form, you'll likely see some mathematical numbers that were generated after processing your data. This makes it hard to track every move of the data you shared, so you must get incentivised in the very first stage of the data exchange.

The advancement in Blockchain technology has allowed for a secure and decentralized way of managing data. By levering this power, data banks can manage identities of the data account owners. Furthermore, it can help track the transfer of data between the verified identities. Blockchain guarantees that the records are theoretically immutable and ensure the privacy of the data. With an added layer of encryption, the data exchange can even be more secure.

Tokens are programmable assets that can be linked to data exchanges. It can represent anything from a store of value to a set of permissions in the physical, digital, and legal world. These tokens, when registered in a crypto exchange, can be exchanged for fiat money. These tokens can facilitate incentivised data exchange with smart contracts that will act as escrows.

In this whitepaper, we propose a data bank for data owners and service providers running on top of a decentralized blockchain network. The Datawallet network is an instance of the virtual data bank where users create their own individual Datawallet for storing their data.

Datawallet's blockchain network is responsible for managing identities and permissions in the virtual data bank. Every transaction is controlled by a smart contract and is logged in a decentralized ledger. Transaction data are secured using Asymmetric cryptography and are only transferred via private channels. Although the channels are private, hashes are shared publicly for verification throughout the network.

Datawallet client is a wrapper that encloses the underlying blockchain network and is responsible for providing an interface for communicating with the blockchain network. It is the client-facing part of the Datawallet that communicates with the network via smart contracts.

The remaining paper is structured in the following manner, Section 2 describes the background literature that will be needed for understanding the architecture, Section 3 lists down the requirements for a data bank and a data exchange. Section 4 derives the architecture of Datawallet from the requirements. Section 5 concludes this whitepaper along with limitations and future opportunities.

In Appendix 2, we talk about various use cases that utilize the potential of Datawallets. Some of them have been implemented whereas others are in active development. In Appendix 3.1, we announce the current implementation of our Datawallet project. In Appendix 3.2, we state the shift to a smart contract architecture and in Appendix 3.3, introduce the marketplace of decentralized apps that will run on the data provided by the underlying virtual data bank.

2. Current Scenario

The internet is a decentralized network with millions of interconnected servers. The more services we consume, the more we feed our data into this enormous amount of servers. Each service manages and operates its own set of servers and has no control over the servers of other services. To use and manage their data, each service also has its data policy and when we agree to consume it, we agree to accept the data policies enforced by it. In this scenario, the data owners and service providers are subjected to three major problems.

The first problem comes on the side of a data owner and it is related to data policies. Companies which provide services may choose not to disclose how they use their users' data. Sometimes they may also not disclose what all information they are collecting. Although their way of collecting and using their users' data might be ethically wrong, the user has already legally agreed to accept their data policies by signing up for their services. Things get even worse when services choose not to abide by their data policies. To summarize, users are not completely aware of the information they are sharing with the services, they don't know if their data is being sold without them knowing or getting incentivized and they also don't know if they are buying back their data, through subscriptions, in some processed format, from some other service to which your data was sold to.

Data traceability is the second problem that arises on the side of a service provider and is related to data authenticity. The data provided by the data owners cannot be trusted and there lacks a mechanism for tracing such information. For example, if a service provider wanted to know your date of birth, he cannot be sure if the data provided by you through the app is authentic unless he gets it verified by some government agency.

The third problem is shared by both data owners and service providers and it is related to data modification. A user might be sharing the same set of data like name, house address, mobile phone number etc with multiple service providers. If the user wants to change his house address he'll have to change it on all the services. Also, it is important for the service providers to keep track of users' data because data is most valuable when it holds the latest version of any information. In the current data ecosystem, the data is sent by the data owners to various services providers at the point of request. Requests done at different points of time can lead to different versions of data. So there lacks a mechanism that can keep the service provider updated with the latest version of the data.

The above problem list is not an exhaustive one but it includes the main problems that have led us to develop a data bank that we call the *Datawallet*, a blockchain-based wallet for facilitating exchanging, tracking and tracing of data, that helps build a better data ecosystem. In subsequent sections, we derive Datawallet's goals and architecture from the above problems.

2.1 Background Literature

(to be done at last)

3. Goals of a Data Bank

In this section, we define the requirements of a data ecosystem at different levels for different entities. These requirements make use of Agent-Oriented Modelling(AOM) with Unified Modeling Language (UML) for making the content more precise and understandable. The ecosystem broadly consists of two main entities, the *Data Owners* and the *Service Provider*. We start by describing the high-level requirements of a Data Wallet, then in section 3.2 we describe the goals of a Data Owner, then finally conclude with the goals of a Service Provider in section 3.3.

3.1 High-Level Goals

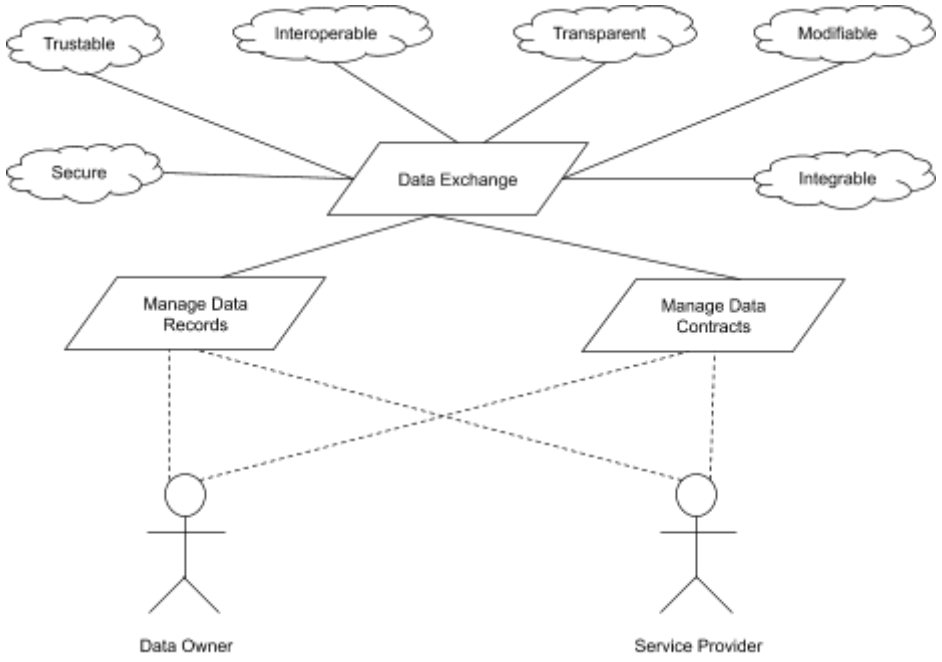

Figure 1: Root goal model for Data Wallet

Figure 1 shows the two main entities of a data bank, like the Datawallet, and the goals they both aim for. These goals mainly focus on data and permissions related to it, which are the most basic requirements of a data bank. To help entities reach their goals the system itself needs to maintain certain quality goals like being secure, trustable, interoperate and transparent.

Secure means that the data bank should be able to avoid unauthorized access and private keys should be offloaded to the end-users device instead of storing it in the data network. *Trustable* means that the entities should be able to trust the identities of each other and should also be able to trust the internal exchange and management of data. Interoperable means that the data bank should be able to operate with external services. Modifiable means that once the data is fed into the data bank, the owners should have the option to update it. Integrable means that external services should be able to directly communicate with the data bank's network for a fully automated process.

3.2 Goals of a Data Owner

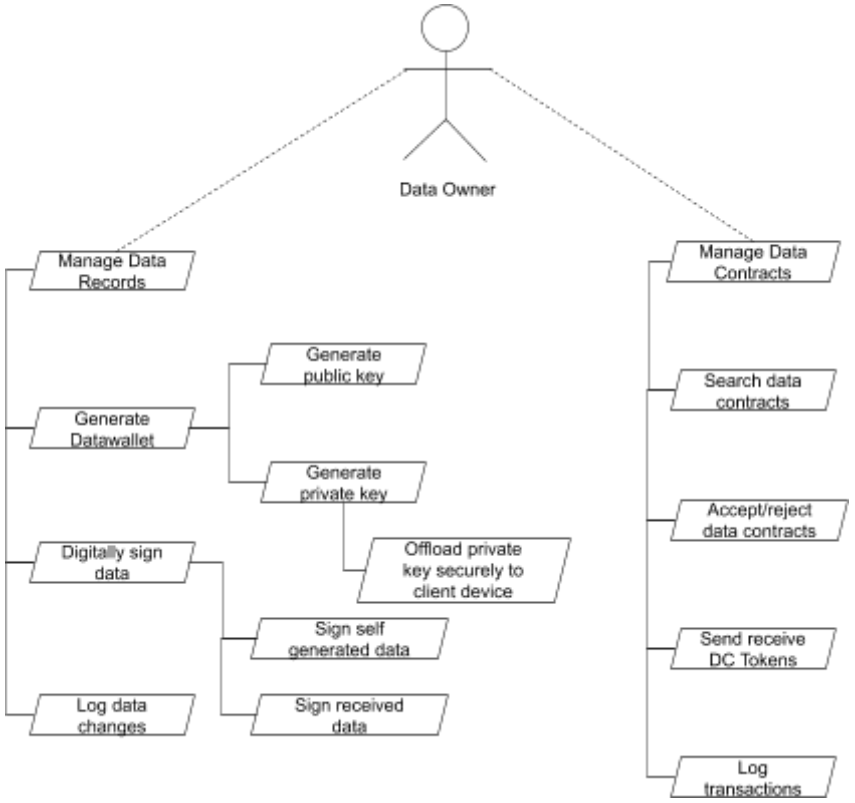

Figure 2: Data management and contract management goals of a Data Owner

3.3 Goals of a Service Provider

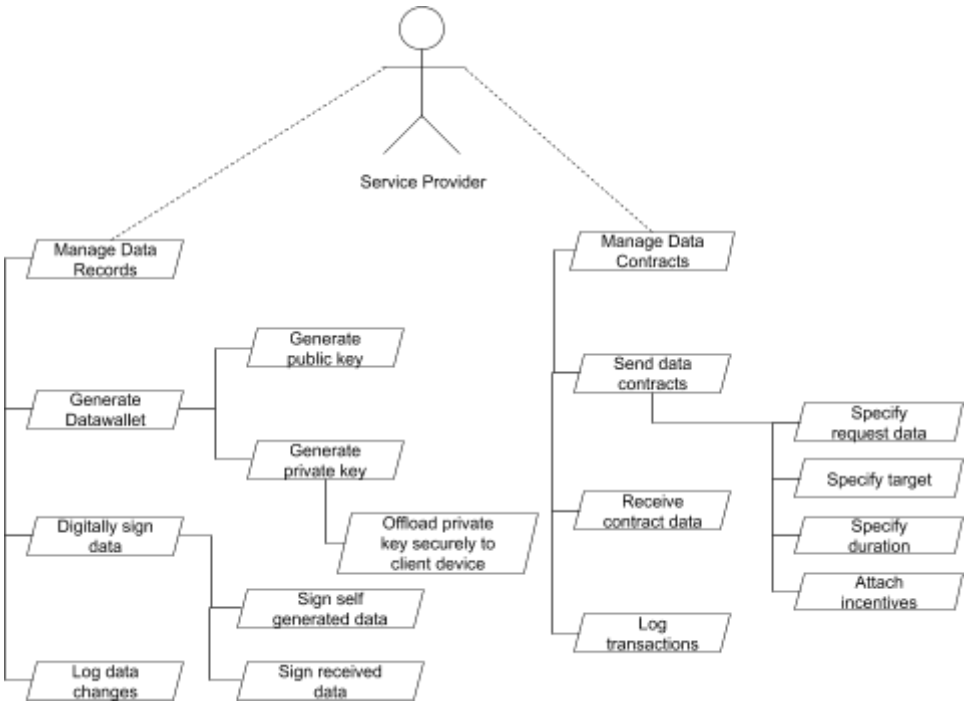

Figure 3: Data management and contract management goals of a Service Provider