

Decentralized network for enriched data exchange

Eric Muysen
eric@hyperbridge.org

Timothy Ko
timothy@hyperbridge.org

Joe Cullen
joe@hyperbridge.org

5th March 2018
Draft v1.0.0

Abstract: Dataforce is a secure smart contract enabled blockchain that will enable users to store, secure and monetize the data that they produce while browsing the web. An implementation of the Republic protocol (see the "[Republic Whitepaper](#)" for details) will be used as the medium of exchange within the Dataforce network to facilitate the selling and buying of stored data, all of which will be done according to the rules and permissions that have been agreed to by our users.

The Dataforce solution is designed for:

II) Private Users: The data economy extracts enormous monetary value from the sale of patterns, predictions and other insights that are gathered from internet user activity, yet the financial benefits are not extended to the people who generate the data. The Dataforce blockchain will provide users the means to securely store their data, and the ability to manage if/how it is collected and monetized. Hyperbridge's digital platform, [Blockhub](#) will serve as the web and blockchain portal in which all data produced by users will have its value securely stored on the Dataforce blockchain.

I) Enterprise Customers: Dataforce will provide the increasingly valuable, and largely untapped, commodity of data to marketers, researchers, and firms that consume data. Enterprise customers will be able to acquire data that has been organized, categorized and prepared for use.

Contents

1 Background	3
2 Solution: a structured private data storage and exchange	5
2.1 Smart Contracts	7
2.1.1 DataStore Smart Contract	7
2.1.2 DataCluster Smart Contract	7
2.2 Data Storage	7
2.2.1 Directory	7
2.2.2 Long-term Storage	8
2.2.3 DataStores	8
2.3 Browser Extension	8
2.4 Token Use	9
3 Security Services	9
3.1 Certification Service	9
4 Long-Term Technical Strategy	9
4.1 Security Auditing	9
4.2 Standards Committee	9
5 Conclusion	9
Acknowledgement	10

1 Background

Early Data

The boom of digital logic circuit industries, and accompanying technologies, such as computers, the internet and digital cellular phones, has resulted in data transforming from a clunky byproduct to one of the economy's most [valuable assets](#).

Following the dawn of the internet there were many restrictions, both technical and otherwise, as to why data could not be fully utilized, and as to why individuals could not manage and secure their personal data. As a means to legally manage data, companies started requiring long and complex service agreements, and most users accept them without reading or understanding the implications. This results in organizations having virtual free reign over the data they collect.

As data continues to grow in importance and value, the implications of centralized control and management of data will become a more relevant concern of both users and businesses. Data has subtly become both a major industry and a dominant resource¹ within modern economies. Used for marketing campaigns, polling, prediction, capturing emerging opportunities, and reaching unserved customer segments, it is now time for the blockchain innovations to reinvent the way in which people manage and control their data and the way in which enterprise customers are able to purchase and access, relevant sections of consumer data. Even the decentralized services being developed at present may collect your data into a centralized silo, thus repeating the mistakes of the past.

“There’s gold in them there mountains of data”²

According to the [IDC](#), “worldwide revenues for big data and business analytics (BDA) will grow from \$130.1 billion in 2016 to more than \$203 billion in 2020”, and the digital universe will grow by a factor of 10 – from 4.4 trillion gigabytes to 44 trillion, more than doubling every two years³. Such a rise in market value and volume begets the “emergence of new tools for collecting and analyzing data and new enterprise roles and responsibilities” concerning how and why data is collected⁴. As this sub-economy develops, questions and concerns regarding the relationship between producers of data (technology users) and collectors/consumers of data (technology companies, marketers, etc.) will come to the fore.

¹ The world's most valuable resource is no longer oil, but data: The Economist, May 6th, 2017

<<https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource>>

² By the end of 2017, revenue growth from information-based products will double the rest of the product/service portfolio for one third of Fortune 500 companies, says IDC. Raw data and various value-added content will be bought and sold either via marketplaces or in bilateral transactions and enterprises will begin to develop methods for valuing their data (see Chief Data Officer below). “Data monetization” will become a major source of revenues, as the world will create 180 zettabytes of data (or 180 trillion gigabytes) in 2025, up from less than 10 zettabytes in 2015, according to IDC.

³ Sisense. “5 Big Data Industry Trends in 2018”. Available at: <https://www.sisense.com/blog/5-big-data-industry-trends-2018/>

⁴ 6 Predictions For The \$203 Billion Big Data Analytics Market, by Gil Press Source:

<https://www.forbes.com/sites/gilpress/2017/01/20/6-predictions-for-the-203-billion-big-data-analytics-market/#ba1eeb208389>

Producers want to ensure their privacy and personal information is being protected, and they want more control over how their data is used, ensuring they are being protected from annoying and invasive digital marketing, and from egregious attacks and from having sensitive information exposed. Meanwhile, consumers of data want to be able to use data more effectively and efficiently, in a means that will be of most impact to their specific aims.

Data's rise [correlates](#) with our capacity to [store information](#) in a permanent way, and has been spurred on by the integration of the general population with networks that gather and share information. The creation of digital records and databases that stockpile and index data, have since enabled data to be commodified in a major way⁵.

Every piece of data a person sends and receives has value, and is monetized and commodified by some platform or another: be it Google, Facebook or Twitter. Vast sums of money are made from every individual's digital footprint, especially so, considering data comes from a growing variety of sources: web browsing, social media posts, devices that aggregate shopper information, photos, videos, purchase records, and GPS signals. Broadening sources of data and an increased capacity to organize and utilize it has resulted in the development of an entire data economy⁶, and as such, data has become the primary resource for increasing profit generation⁷.

Collateral Damage

During the process of data collection, personal privacy and security can be compromised resulting in real world consequences⁸, and as it stands internet users do not possess adequate tools to secure their personal information and data and to protect themselves from various real world and cyber threats⁹

Right now most internet services and databases are centralized, collecting data into large, unmanageable silos. In time, these silos are auctioned off to the highest bidder, but most of the data is not fully-utilized, and as a result large quantities of usable data goes wasted. The potential to structure, manage, and efficiently monetize this data is going untapped.

⁵ E. Schoenherr, Steven (5 May 2004). "The Digital Revolution"

⁶ Elements of the European data economy strategy
<https://ec.europa.eu/digital-single-market/en/towards-thriving-data-driven-economy>

⁷ "From a geographic perspective, more than half of all big data and business analytics revenues will come from the United States. By 2020, IDC forecasts that the U.S. market for big data and business analytics solutions will reach more than \$95 billion."

Source: <https://www.idc.com/getdoc.jsp?containerId=prUS41826116>

⁸ FTC Recommends Congress Require the Data Broker Industry to be More Transparent and Give Consumers Greater Control Over Their Personal Information: Agency Report Shows Data Brokers Collect and Store Billions of Data Elements Covering Nearly Every U.S.

Consumer <http://bit.ly/1Hc8yGy>

⁹ "[COMMENTS SUBMITTED TO THE VERMONT DATA BROKER LEGISLATION WORKING GROUP](#)"

Not all data is created equal

Although data can provide valuable insights into the challenges faced by all facets of our society and economy, it needs to be of a certain quality and structure prior to becoming an effective tool. Although progress has been made, much of the latent potential of bulk data produced goes wasted. Partially because it remains unstructured (no predefined data model). There remains an urgency for technology firms to innovate and develop solutions to process data and transform it into something that is more easily collected, interpreted, and utilized.

Unstructured Data

Much data remains in an unstructured state, and is useless for companies since in this state it cannot be organized or used to provide any valuable insights. Accessible raw data needs to be transformed into a usable and scalable format before it is able to be used or monetized. The transformation from raw data to ripe data needs a lot of effort and process. This transformation will create several data clusters that would help the companies to do analysis or get insight for their business. For example, the data that has been transformed will be able to cluster the consumer data into several categories such as socio-demographic profile, social network and relations, online preferences, purchasing preferences, channel preferences, location, wealth and even risk level.

2 Solution: a structured private data storage and exchange

Dataforce Network will be a suite of smart contracts on the Ethereum network that will facilitate the exchange of data in a secure and granular way. The network is comprised of users producing data (“Data Producers”) and consumers of data (“Data Consumers”), such as enterprise clients and marketeers. Data Producers are any software or hardware that track and push data into the integrated Dataforce network. An example of Dataforce integrated software is the Blockhub Platform, an app marketplace that facilitates the collection of data within the Dataforce Network (Please see the [“Blockhub Whitepaper”](#) for further details). Other integrations we envision include Apple HealthKit, Stava, Spotify, TicketMaster, Facebook, Twitter, and more. When these services are bridged with Dataforce, Data Producers will be able to encrypt their data into any supported decentralized database and release, or sell their data in a granular means based on parameters that they have pre-determined. Data Producers will then have the option of being compensated when a Data Consumer’s data collection filter correlates with their available data.

Dataforce will create a data enrichment standard that will meaningfully label and categorize both historic and continuously generated user data. By implementing a categorization standard and a process for migrating user data to encrypted and distributed ledgers, control of data will transition from centralized organizations to the decentralized web of data creators.

Enterprise Benefits

Due to the enhanced richness of data organization and maintenance, this model will activate and provide value to swaths of data that are either wasted or under-utilized. Dataforce will provide usability and access to aspects of the data realm that have gone untapped due to various architectural and technological limitations (addressed in Section 1.2).

User Benefits

In addition, Dataforce is designed to allay concerns of privacy that people have over the control and management of their data. The data we collect will be encrypted, adding the benefit of security. User-identifying data of a sensitive nature can optionally be stored separately from marketer-relevant demographic data that will be put on to the blockchain.

Dataforce will integrate applications that manage access to data, ensuring that if data is collected and transformed into scalable and structured data, that originators of the data will have the option and ability to monetize it.

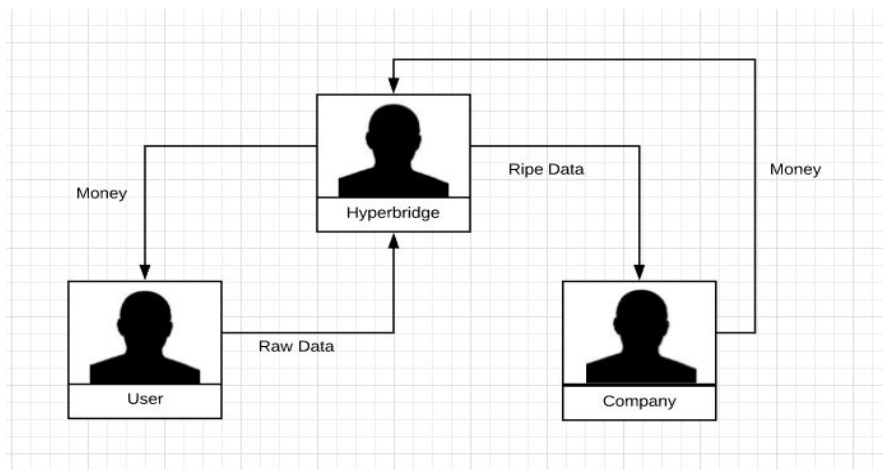


Figure 1. Dataforce data monetization scheme.

Dataforce will use multiple components as part of its own blockchain ecosystem. Directory service will be provided using a smart contract based blockchain, which will initially be a fork of the Ethereum blockchain. Users will be able to claim the initially issued ERC20 for a token on the official blockchain.

Dataforce will use the Republic Token, a [ERC20 standard token](#) with decentralized governance built into it. For more information on the Republic Network see the [official website](#).

Dataforce solves trust, and complicated data transformation issues while allowing users to manage how their data is accessed and used. This ensures that if data is collected, it is not done surreptitiously, and that the benefits and compensation are reaped by the originator of the data. By

creating a standardized and encrypted data structuring model, Dataforce is enabling a more valuable and direct marketplace for marketers to select marketing targets.

2.1 Smart Contracts

There will be a number of contracts in Dataforce to facilitate the storage and access of data operations on the network, both historically and continuously generated.

- DataStore Smart Contract
- DataCluster Smart Contract

2.1.1 DataStore Smart Contract

Dataforce will use a smart contract for the organization of user data. When a user synchronizes their data, Dataforce will encrypt their sensitive data and insert it into the decentralized database. It will then execute this Ethereum contract providing a reference to that data blob and some metadata (for organization and searching). Sensitive user data will then be removed from their local machine. During synchronization, this contract will be checked for new changes. During migration, the full history will be downloaded from this contract and used to collect data for migration to the destination blockchain.

2.1.2 DataCluster Smart Contract

Dataforce will use a smart contract to access structured datastores using data points, such as category of data, demographic of user, etc. These clusters are high-level representations of the raw data gathered by users and made more accessible for organizations to derive useful results.

2.2 Data Storage

2.2.1 Directory

A directory comprised of high-level data metrics for ease of filtering and locating search criteria matches.

2.2.2 Long-term Storage

Dataforce will utilize distributed database technology for long-term storage of encrypted user data. Initially we look to support IPFS, with other options as they arise and become economical.

2.2.3 DataStores

DataStores are raw data containers encrypted by the user and put into long-term storage. Although the datastores can be encrypted any way, we recommend encrypting with a 2-factor device, such as a physical hardware wallet, such as Trezor or Ledger. An initial goal will be to provide a library and documentation to do this with a Trezor.

2.3 Browser Extension

Dataforce will be an extension that pipes data to supported clients, such as the Blockhub Platform. However, it does not do any processing itself.

Below is a diagram showing how the browser will communicate with a supported client running in the background.

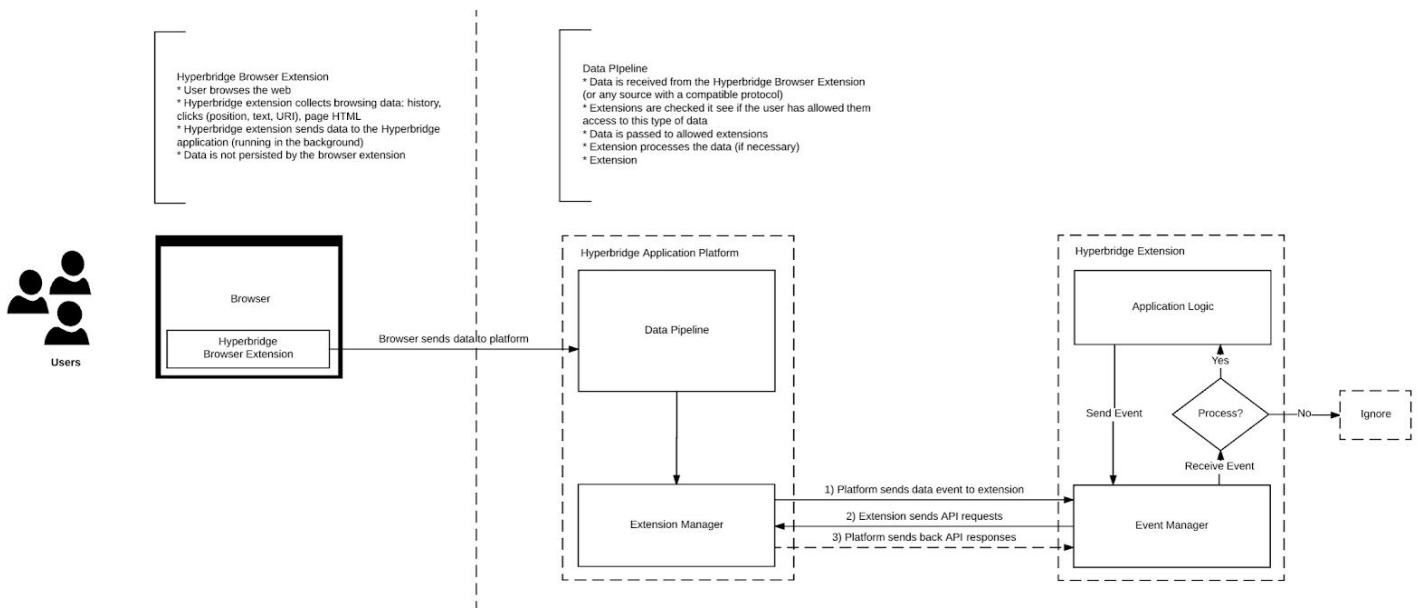


Figure 9: Utilization of data gathered during a user browsing experience.

2.4 Token Use

An instance of a Republic Token will have various uses within the Dataforce data exchange. The Republic Token will be required for specific operations within the ecosystem, including:

- Data providers will be compensated for data with Republic tokens.
- Data consumers will pay for data with Republic tokens.
- Data operators will accept Republic tokens for their services.

3 Security Services

3.1 Certification Service

Hyperbridge Technology Inc. will offer certification for entities interested in bridging with the Dataforce Network. Certification will give users confidence in the authenticity and security of the provider's app. Contact partners@hyperbridge.org for more information on this service.

4 Long-Term Technical Strategy

4.1 Security Auditing

We will undergo regular security audits for the Dataforce software and smart contracts.

4.2 Standards Committee

A committee of the leaders in the blockchain and related technology space will be formed to oversee the progress of the open source specifications. We will work closely with regulatory and standard bodies, including but not limited to ISO, IEC, and W3C.

5 Conclusion

Dataforce is posing both user-centered and enterprise solutions to the current problems faced by big data. Dataforce will address security and management concerns, providing users the ability to directly assign value and monetization potential to the data that they create, allowing them to sell their data, whole or piecemeal, to data consumers.

Dataforce also seeks to address the inefficiencies and lack of organization and utilization present within the present model of data collection and use, ensuring that the originators of data (users) maintain the control and capacity to manage this increasingly valuable resource.

Dataforce will provide marketers, researchers, charities and the like with the most efficient and frictionless access to user-generated data, thus creating a fair and transparent market for the collection and use of personal data, all the while respecting our users' terms and conditions.

Acknowledgement

We would like to thank David Mayoh and Walid Al Habboul for contributing to, editing and formatting this paper.