Abstract

Consumers (Data Providers) do not participate in the profits generated from the data they create. Consumer personal data fuels hundreds of billions of dollars in revenue per year across many industries. The Data-Driven Marketing Economy (DDME) was most recently estimated, in 2014, to be a $200bn+/year business in the United States alone [1]. The DDME figure does not include the revenue-generating role data plays in such industries as credit and insurance underwriting, content-provision such as streaming services, or e-commerce provision. Personal data include digitally logged static attributes like age and gender, but also dynamic data like location, spending history and search and browsing habits, among others. Businesses in all economic verticals derive value from trends and patterns gleaned from analysis of aggregate consumer data. Additionally, businesses with direct, one-to-one, consumer relationships derive value from individual-level data exchange. But faulty Internet infrastructure, and all the centralized consumer-layer platforms built on top (e.g.; Social media like Facebook and search engines like Google) exploit consumer data for their own profit. Thus the key problem: consumers do not "own" or control their digital data and have no means of monetizing it. By individually signing up for digital platforms and services, consumers collectively create a mass data resource that can be exploited repeatedly, for economic benefit, by both the businesses that host and control data and the customers with which they trade it. The solution is a decentralized data marketplace for consumers to control and exchange personal data directly with businesses as they see fit. Consumers would gain from monetizing their data and businesses would gain from accessing more accurate, consent-based data. To do this, we introduce the personal data token (PDT) ecosystem.

Glossary of Terms:

Aggregate (adjective): Formed or calculated by the combination of separate units. For example, in our ecosystem an aggregate data set would be formed from combining data records from many unique data providers.

Computation: Computation is any type of calculation that includes both arithmetical and non-arithmetical steps and follows a well-defined model. For example, we may want to derive a sample size from an array based upon specific conditions, i.e.; the number of 25-35 year old males in an array.

Decentralized: To decentralize is to move organizational or administrative power away from a Central actor in a system. In the case of our white paper, we use the term to describe a state in which there is no central system power to censor or change rules unilaterally.

Ecosystem Administrator: The administrator sets basic guidelines for transactions in the marketplace; provision of initial technologies and other administrative objectives like allocating budget spend.

Layer 2: Decentralized networks, or protocol solutions that operate “on top” of existing blockchains. Computation is moved off-chain, either to enable privacy or to save computing resources. Often times, layer 2 solutions will operate their own consensus network, and anchor computation results to an underlying blockchain, so that the system maintains the blockchain as the source of truth.

Protocol: A defined set of rules or standards by which actors or objects in a system are bound.

Self-Sovereign Identity (SSI): Identity is the persistent fact of being who or what a person or thing is. SSI is the notion that who or what a person or thing is, is defined by the person or thing, and not assigned to the person or thing. For example, in our ecosystem, the data provider declares that they own the device with which they register for the ecosystem. Control of that device is the origin fact that will be the nucleus of all other claims made by the data provider.

Trusted Execution Environment (TEE): TEE is an isolated area on the main processor of a device that is separate from the main operating system. It ensures that data is stored, processed and protected in a trusted environment. It is used in our ecosystem to preserve privacy and decentralization while computing on data.

Contents

1 Introduction to Personal Data

1.1 Current Architecture: Consumer-Facing Platforms

Digital identities are assigned to consumers by platforms like Google, Facebook, Amazon, JP Morgan and most other businesses with a digital presence. The consumer data associated with digital identities are stored in centralized databases that are controlled by these businesses. This arrangement makes the business the *de facto* owner of the data. This simple, yet critical architectural underpinning provides for businesses to generate revenues and profits from selling, or otherwise monetizing, consumer data, without remuneration to the consumer. If consumers controlled access, and were *de facto* owners of their own data, then:

- Businesses would not be free to repeatedly exploit data solely for their economic benefit.
- Consumers would share in the benefits reaped from data sales, access provision and other business usage.
- Businesses would have access to more accurate, consent-based data from sources outside of their specific data purview.

The pie of data value is not fixed. A consent-based model, including more accurate and holistic data, grows the pie for both sides to benefit.

1.2 Current Architecture: Data Brokerage and 3rd Party Platforms

Adjacent to 1st party, consumer-facing platforms like Facebook, 3rd parties, like data brokers and various other data intermediaries (known broadly as the DDME), generate revenue by buying and selling consumer data that is either made available by 1st party platforms, or tracked and compiled surreptitiously by the 3rd party brokers themselves. Unlike 1st parties, such as social media, or banks, 3rd party data brokers and intermediaries have no relationship with the consumer whatsoever, so their exploitation of data is more egregious than a 1st party's. In exchange for permitting data access to JP Morgan or Facebook (1st parties), a consumer receives apps, products and services in return. That the 1st party has a direct relationship with the consumer tempers the incentive to unabashedly exploit the consumer's data. 3rd parties - like the data broker Acxiom - on the other hand, have no exploitation constraints. They do not have a relationship with the consumer and thus will exploit consumer data, for profits, to the highest possible degree. That the digital economy has evolved to its current untenable form is indicative of its structurally centralized underpinnings.

1.3 The Fatal Flaw of Centralization

As outlined in the Abstract, centralized platforms that host data decide when and how to use, or exchange it and accrue almost all of the economic benefits. However, they bear little cost when something negative happens (hack, breach, selling to a dubious counterparty) - costs that accrue, almost exclusively, to the consumer. Centralized databases present a rich target for nefarious actors. The 2017 hack of Equifax's database [2] is a relevant case, in which consumer data was hacked from centralized storage owned and controlled by Equifax. Monetary costs related to identity theft, including insuring against future mishaps, and other preventative measures are borne by the consumer. The irredeemable cost is that sensitive data is forever leaked to bad actors, the effects of which can materialize years after a breach. Meanwhile, Equifax generated $3.4bn in 2017 revenue [3], from storing, buying and selling consumer data. This is textbook moral hazard. They participated in all of the economic upside of data value, and the consumer is left to shoulder the costs – the downside and after-effects of hacks and breaches – associated with faulty centralized architecture.

1.4 Monetizing Data: Existing and New Opportunities

Our primary functional objective is to help consumers get compensated for making their data available to businesses for commercial use. With adoption of data control, the consumer becomes the point of data distribution (the data provider), thereby supplanting the role (and re-appropriating the economic benefits) of the centralized business or platform that previously controlled the database. With increased data agency, and a marketplace outlet, the data provider will find expansionary ways of utilizing and ascertaining value from distributing their data. Some promising use-cases, like the Hancock insurance - Fitbit partnership [4], portend a world in which a data provider can insert data into many different transactions to generate increased value. In this particular case, Hancock insurance offers lower premiums on life insurance policies to prospective policyholders that share Fitbit data. The insurer gains by receiving a more accurate and recent version of relevant data. The data provider gains by lowering their premiums. This type of data transaction could potentially span across many different economic verticals including lending, telecom, retail and more. Having a current user-base of data providers (and thus data), Datacoup has seen significant interest and demand for more holistic, more accurate and more recent, consent-based data from market participants named above, in addition to participants like market researchers and investment analysts.

A new paradigm of consumer data ownership will both facilitate a massive economic shift of existing data-driven revenue streams (from businesses to consumers), and also open expansionary economic opportunities for consumers to utilize and monetize their data asset. Such opportunities stretch farther than monetization alone, and include the provision of data to app developers for new and more personalized services. We'll describe how we can fulfill the promise offered by a world in which consumers own their data.

2 PDT Ecosystem Overview

2.1 Blockchain Technology and Personal Data

In 2008, Bitcoin was conceptualized as the first decentralized digital currency [5]. In the infamous white paper, Satoshi Nakamoto designed a peer-to-peer digital currency system that sought to address the underlying issues with single institution/entity control of money. The manifestation of the concept outlined in the white paper gave us a cryptographically secure digital asset (BTC) that can be owned and transacted, by network participants, without having to trust single entities to keep a ledger of ownership records and transactions.

Analogous to money, personal data is plagued by single points of trust. As described in section one, centralized entities abuse power and use their status as de facto data owners to reap economic benefits, without bearing the cost of negative outcomes stemming from faulty centralized architecture.

The properties inherent to decentralization – increased security, and 'trustlessness' – make blockchain architecture attractive for consumer data-ownership.

Like money, data requires cryptographic security and a ledger of ownership in which transactions are processed and records are maintained by an open, permissionless network. This obviates the need to trust single entities that can either be corrupted, abusive or act as a central point of failure.

2.2 The PDT Ecosystem

The PDT ecosystem is a blockchain-based marketplace for data providers and data requesters to transact consumer personal data.

Here is an ecosystem diagram outlining the main actors, objects and resources in the ecosystem:

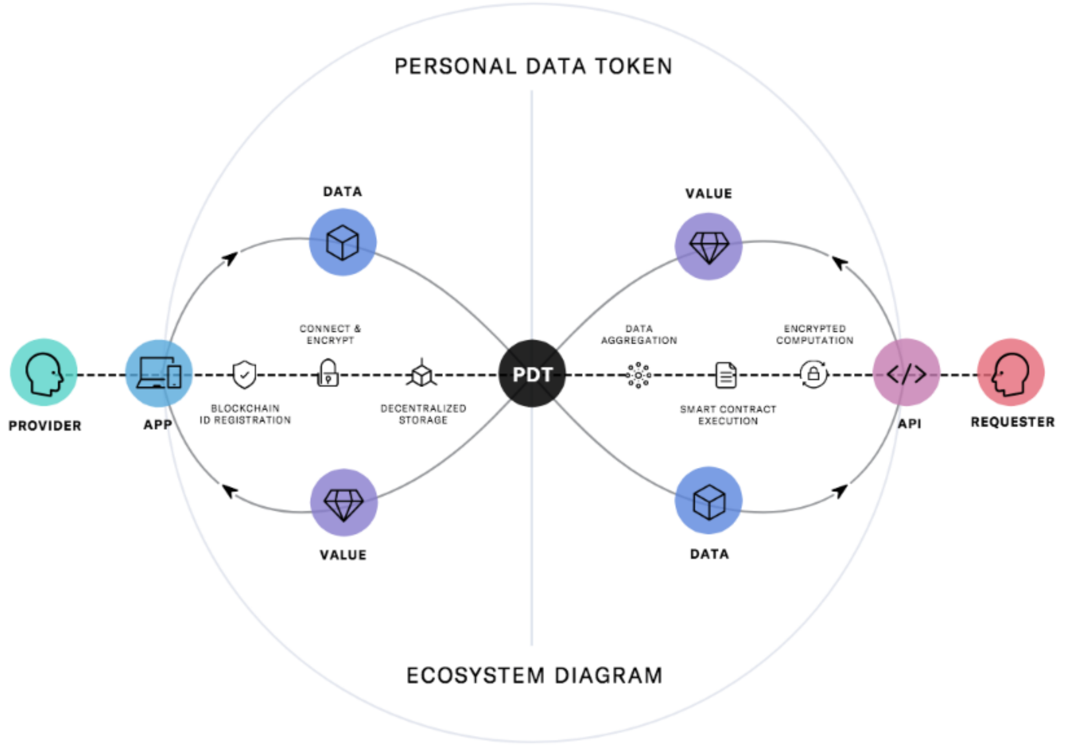


Figure 1: Overview of the PDT Ecosystem

The key ecosystem resource is consumer data, including spending, search and browsing, location, social, service consumption, demographic, fitness, IoT and other data sources. Data providers connect data and make it available for sale or access in the marketplace, in exchange for tokens or other value, like products and services.

As in any marketplace, or network, as more of the key resource is made available for exchange, more participants and capital are drawn to the ecosystem. This is the classic flywheel effect of marketplaces. More value can be created both from higher numbers of data providers willing to exchange their data individually, and from a larger aggregate data-set, from which businesses can derive value. With data, in particular, there is also expansionary value created as the overall amount on the platform scales and consumers/businesses propagate new ways of unlocking value from the resource. Thus, the PDT ecosystem would become more valuable as more data providers connect and make data available for exchange. Compared to other similar projects, we have significantly more experience with both sides of the market – Data Provider and Requester – that will prove advantageous when scaling the marketplace.

2.3 Ecosystem Participants

Participants Summary

Data Providers - Known colloquially as "users", data providers are individual consumers that create data and make it accessible to data requesters in the marketplace. Across existing Datacoup apps, we have acquired a 5-digit user-base of data providers who've connected approximately 60,000 different data accounts.

Data Requesters - Any entity in the ecosystem that requests access to data, including

- Requester – Aggregate Data (RAD): A purchaser of aggregate data that is seeking to understand consumer trends and patterns
- Requester - Individual Data (RID): A vendor that wishes to access or purchase data from an individual customer, usually for marketing or CRM purposes
- Requester – Developer (RD): App developers building on robust data to provide consumers with apps or other ecosystem services

Ecosystem administrator – The administrator sets basic guidelines for transactions in the marketplace, provision of initial technologies and other administrative objectives. Datacoup will serve as the initial ecosystem administrator.

Participants Detailed

Data Providers

Individual consumers (data providers) create and make available the key resource to the ecosystem: Data. Across existing Datacoup apps, we have acquired a 5-digit user-base of data providers who've connected approximately 60,000 different data accounts. Data providers will be compensated with PDT for connecting specific data sets via available Datacoup apps. PDT allocated for this purpose is held in treasury until release.

The ecosystem administrator is tasked with PDT distribution from Treasury. We anticipate the lion's share of treasury tokens will be distributed to data providers for connecting data and making it available for sale, or access. The administrator will set initial token compensation rates for data connections.

As the marketplace matures with ample supply and demand, prevailing market rates will dictate PDT remuneration to data providers.

Data Requesters

Requester – Aggregate Data (RAD)
RADs pay fiat to access data in the marketplace. Like other requesters, RADs must purchase and maintain token stakes in order to access the data marketplace. For ongoing purchases, RADs will pay fiat that will be converted to PDT and distributed to data providers whose data is in the purchased set. Typically, RADs are interested in accessing aggregate data sets or computed outcomes from aggregate data sets. This is distinct from entities or businesses that want to market products or services to data providers.

RADs include entities such as investment managers, market research companies, brands and others. RADs will have access to a rapidly growing consumer data set, encompassing disparate data from persistently fed API sources. They will have the ability to sort data based upon such inputs as demographics, consumer purchases, searched items, cart abandonment rates, location and other key data inputs, toward the end goal of analyzing consumer behavior, establishing correlations, and generally finding patterns that will add value to their business processes.

Requester – Individual Data (RID)
RIDs are comprised of marketers and other vendors that wish to market goods or services to data providers in the ecosystem. RIDs, like all requesters, must purchase and maintain token stakes in order to access the data marketplace. Such partners might be banks, insurance companies, retailers, telecoms, content providers and more. Typically, a RID is trying to increase sales of a product or service and will be soliciting data providers with coupons or discounts, via data targeting, to do so.

Once use-case might be a bank, with a pre-existing depositor relationship, incentivizing the depositor to pledge more data to the bank in return for waived checking account fees.

Requester – Developer (RD)
RDs utilize the data provider's data to build applications or provide ecosystem services. Given the depth, breadth and scope of the underlying data resource being built in the PDT ecosystem, RDs have an opportunity to build apps and other services on top of an extremely rich and robust data-set. Presumably, the best AI will be trained on the best data.

In the early stages of the ecosystem, RDs may receive token grants or "seeds", as incentives to build apps and services for data providers. At scale, however, we anticipate that RDs may find the specter of selling apps/services to a large user-base to be enough incentive to build on top of the data. There is a case to be made that in the future, RDs stake PDT, like other data requesters, for ecosystem access to build apps. We anticipate data providers to pay for apps with PDT earned from data sales. In return data providers can expect rich app experiences, recommendations and other premium AI/ML services.

Ecosystem Administrator

As creators of the ecosystem, the Datacoup team will be the initial ecosystem administrator. The ecosystem administrator is tasked with charting the ecosystem on a course toward maximizing token value for all participants involved. This includes steering technology updates, identity and reputation implementations, standards for data transactions and the roadmap toward decentralized governance. Initially, this role will resemble that of the NEO council outlined in the NEO project [7]. However, as implied by "roadmap to decentralized governance", the ecosystem will be moving toward a decentralized system. The ecosystem administrator role will decay over time and be filled by a sound governance structure that funds treasury for development needs, and provides a voting mechanism for how budget is to be spent.

3 Technology Overview and Layers

3.1 Technical Overview of the System

Before we breakdown, in detail, the key protocol components used in the technology stack, we want to offer a technical system overview. This provides for familiarity with the core supply and demand actors in the marketplace and how they interact with key technology – apps and components – in the system:

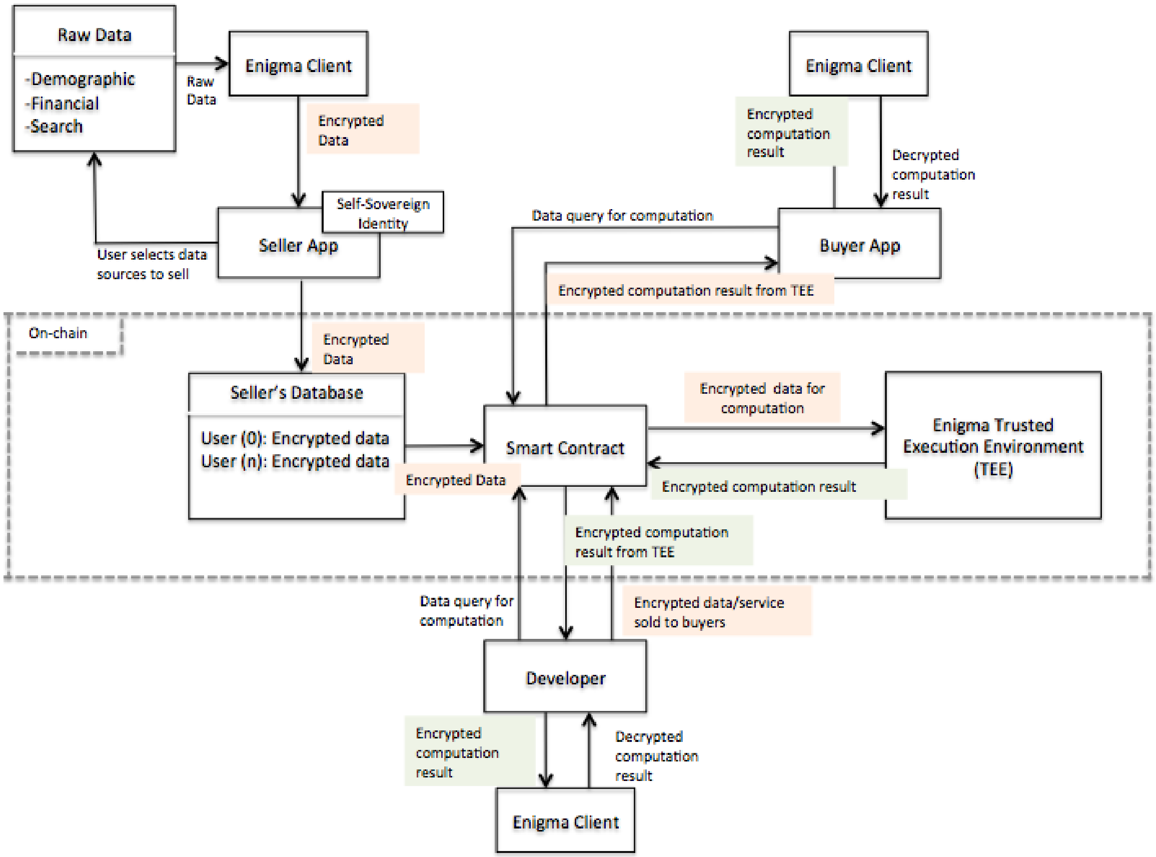

Raw Data
-Demographic
-Financial
-Search
Raw Data
Enigma Client
Encrypted Data
Self-Sovereign Identity
Seller App
User selects data sources to sell
Enigma Client
Encrypted computation result
Decrypted computation result
Data query for computation
Buyer App
Encrypted computation result from TEE
On-chain
Encrypted Data
Seller's Database
User (0): Encrypted data
User (n): Encrypted data
Encrypted Data
Smart Contract
Encrypted data for computation
Enigma Trusted Execution Environment (TEE)
Encrypted computation result
Encrypted computation result from TEE
Data query for computation
Encrypted data/service sold to buyers
Developer
Encrypted computation result
Decrypted computation result
Enigma Client

Figure 2: Technical Overview of the PDT Ecosystem

3.2 Simplified Technology Stack

Below describe the technology layers in the decentralized tech scheme, agnostic to the specific technology used in Datacoup apps.

- Blockchain - The blockchain serves as the base layer source of truth for data ownership and transactions. In centralized architecture, all key app transactions and state changes would be logged in a database on a platform-owned server for consistency and availability. Instead, these functions now resolve to the blockchain.
- 2nd layer protocols - On-chain functionality is limited and has led to the growth of myriad off-chain, network-based solutions and protocols to handle key functions like compute and store. We refer to these protocols as layer 2 solutions.
- Applications - On top of the 2nd layer functions are apps/dapps. Apps are the User Interface (UI) for app users to access blockchain-based networks and services. Apps rely on 2nd layer protocols and the blockchain for trustless execution and record keeping of data transactions. Along with the 2nd layer, apps abstract direct communications with the blockchain to deliver users a more frictionless experience.

Here is a diagram of the simplified technology stack:

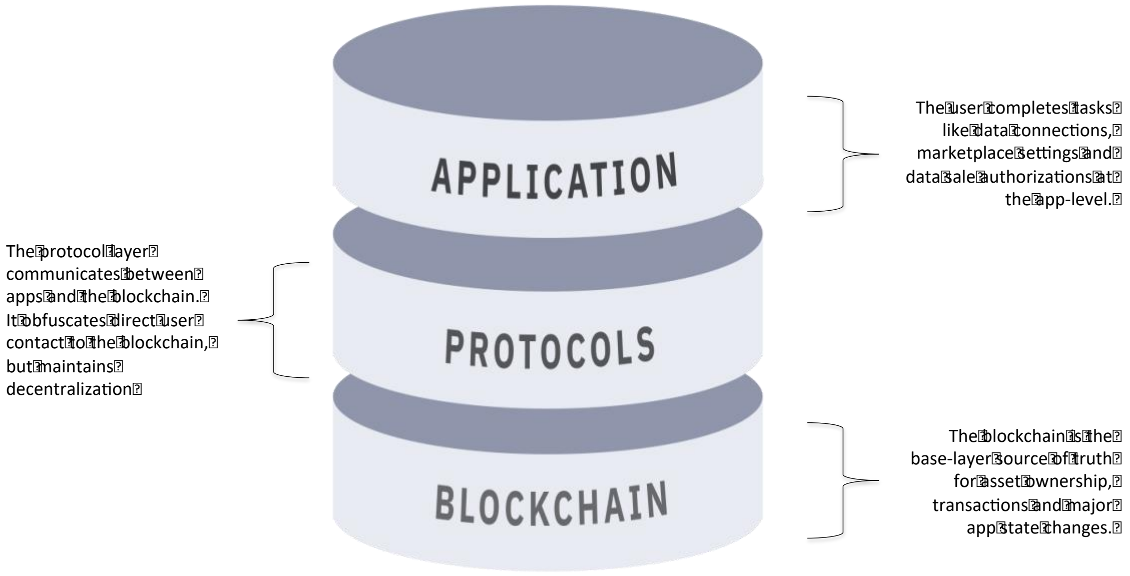

APPLICATION
PROTOCOLS
BLOCKCHAIN
The user completes tasks like data connections, marketplace settings and data sale authorizations at the app-level.
The protocol layer communicates between apps and the blockchain. It obfuscates direct user contact to the blockchain, but maintains decentralization
The blockchain is the base-layer source of truth for asset ownership, transactions and major app state changes.

Figure 3: Technology Stack - Simplified View

Considering the above as agnostic, below we'll describe the apps, protocols and blockchain technology layers specific to the PDT ecosystem.

3.3 PDT-Specific Technology Stack

Datacoup Apps

Data providers interface with and transact in the decentralized data marketplace via the Datacoup app-suite. The initial suite will include a web-app, browser extension and mobile app. The full app-suite offers the data provider myriad ways to connect valuable and disparate data sources, and to exchange their data for remuneration as they see fit. Having acquired data providers across 5 different web-apps, our experience gives us a distinct advantage in understanding data provider behavior related to normative expectations involving data-set connections, compensation requirements and more.

- *Web-app*: The web-app is the central hub for functionality in the app suite: Data connections, marketplace preferences, account balances and transaction histories are all available.
- *Browser Extension*: The Browser extension is a key tool for a data provider to capture their valuable search and browsing data. Because functionality stems from the browser, the data provider can also permit and transact data on any website. The data provider can connect a more abbreviated set of data sources, and also will have access to all account balances and some portions of historical transactions.
- *Mobile app*: The mobile app is the source for a data provider's captured location data. It also offers functionality for the data provider to connect and exchange data anywhere and anytime. The Data provider can connect a more abbreviated set of data sources and also will have access to all account balances and historical transactions.

Layer 2 Decentralized Protocols

The Datacoup apps utilize protocols that cover 3 key areas that conduct both off and on-chain core functions:

1) An attestation model for Identity and registration
2) Decentralized data storage
3) Private and decentralized compute and transact

The Datacoup apps, and the protocols they leverage, write to the blockchain to record critical items like asset ownership (data, identity or tokens) and transactions involving those assets. The blockchain is the underlying source of truth for digital asset ownership, asset transfers and major state changes for apps in the ecosystem. However, scaling issues with on-chain functions [6] have engendered a bevy of off-chain, or layer 2, solutions for many of these functions. Dapp UX's are extremely strained when every in-app task requires a write to the blockchain and pursuant payment in the native chain's currency. For these reasons we're using off-chain solutions for the above-mentioned critical tasks.

Blockchain

The PDT ecosystem is using the [____] blockchain for marketplace smart contracts. It is the most Scalable and the most advanced blockchain from a capability standpoint.

Data transactions in our marketplace require automation. For example, when a requester purchases data, the smart contract should initiate the process of pulling data from decentralized storage into the TEE for computation, and sending escrowed tokens from the contract to the data provider, for selling data.

Building from our simple diagram above in Figure 3, this diagram shows the technology stack specific to the PDT ecosystem:

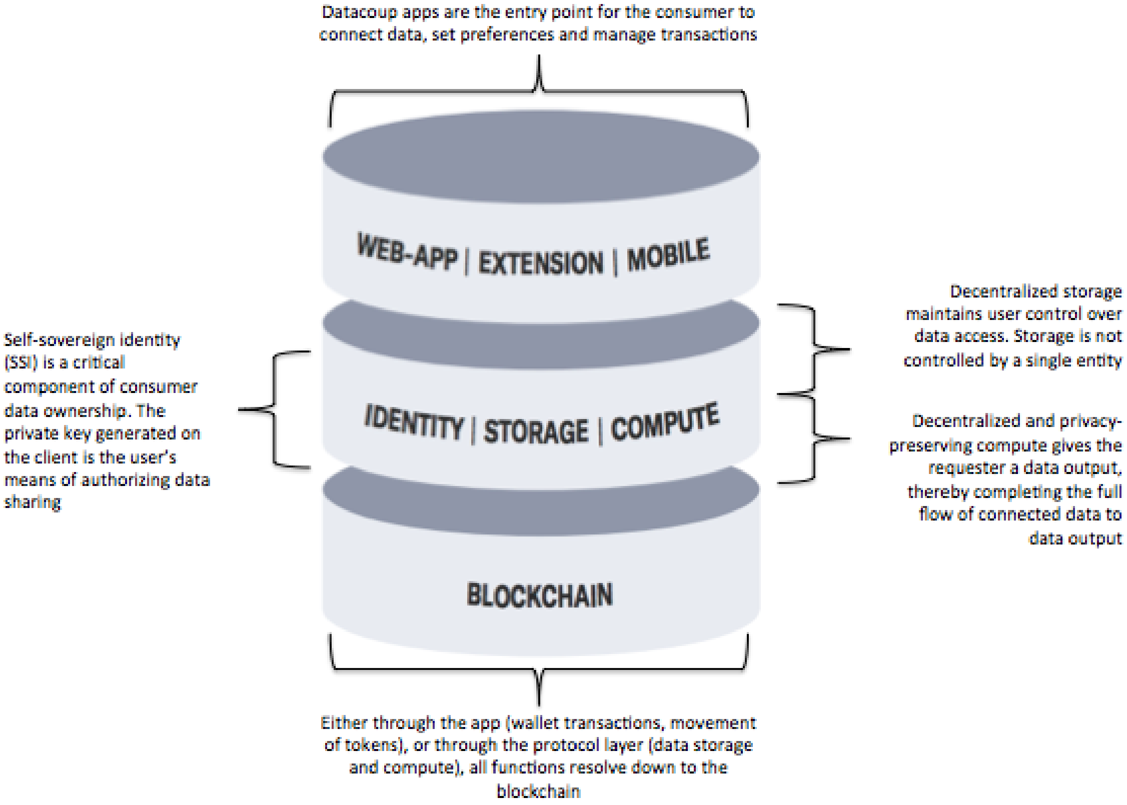

Datacoup apps are the entry point for the consumer to connect data, set preferences and manage transactions
WEB-APP | EXTENSION | MOBILE
Self-sovereign identity (SSI) is a critical component of consumer data ownership. The private key generated on the client is the user's means of authorizing data sharing
IDENTITY | STORAGE | COMPUTE
Decentralized storage maintains user control over data access. Storage is not controlled by a single entity
Decentralized and privacy-preserving compute gives the requester a data output, thereby completing the full flow of connected data to data output
BLOCKCHAIN
Either through the app (wallet transactions, movement of tokens), or through the protocol layer (data storage and compute), all functions resolve down to the blockchain

Figure 4: Technology Stack - PDT-Specific View

4 The Components: Layer 2 Protocols

In the above section 3.3, we outlined three key areas in which protocol components must interact with each other, Datacoup apps, and the blockchain. We'll revisit those concepts here, in greater detail, in order to shine a light on how we accomplish a privacy-preserving method for a data provider to exchange data with a data requester.

To explain the protocols and their interactions with Datacoup apps and the blockchain, we introduce a case that will guide the explanations for the three major protocol layer components.

Case: A data provider joins the PDT ecosystem via the Datacoup web-app, in order to monetize data. The data provider creates an identity, connects data sets and makes *data available for exchange. A hedge fund data requester sets requirements to purchase access to an aggregate data set in the marketplace. The data provider's data is pulled from storage into a secure enclave, among other provider data, in order to compute the particular data output stipulated by the hedge fund requester. Upon receipt of the data output, a data provider is remunerated for their participation in the transaction.*

4.1 Component 1: Identity

A critical first step for the data provider to transact in the data marketplace is identity creation in the system. To every degree possible, the goal of the identity system is to validate the integrity of the data provider as a unique, and legitimate source of data. This important step assures that data requesters can confidently purchase data from verified, real data providers. Without instilling the assurance of data integrity, the marketplace would not be possible. We utilize an attestation model that verifies a data provider's control of the device, a phone number and an email account. Once that has been established, a public/private key-pair is generated on the verified device. The private key will serve as the data provider's authorization mechanism for future marketplace actions like adding data sets, permitting data to requesters and more. Necessarily, ownership of the ID must be expounded on the blockchain, so that all other exchange participants can be confident in transactions with this counterparty. This is not to say that identity-related personal data is written to the blockchain. On the contrary, what are being written to the blockchain are event occurrences that signify the origin record of a data provider identity creation in the system:

- Data provider device-control and phone/email authentication has been verified at a particular time
- Public/Private key has been generated, locally on the device and secured

Here is a sequence diagram exhibiting how component #1 accomplishes its goal.

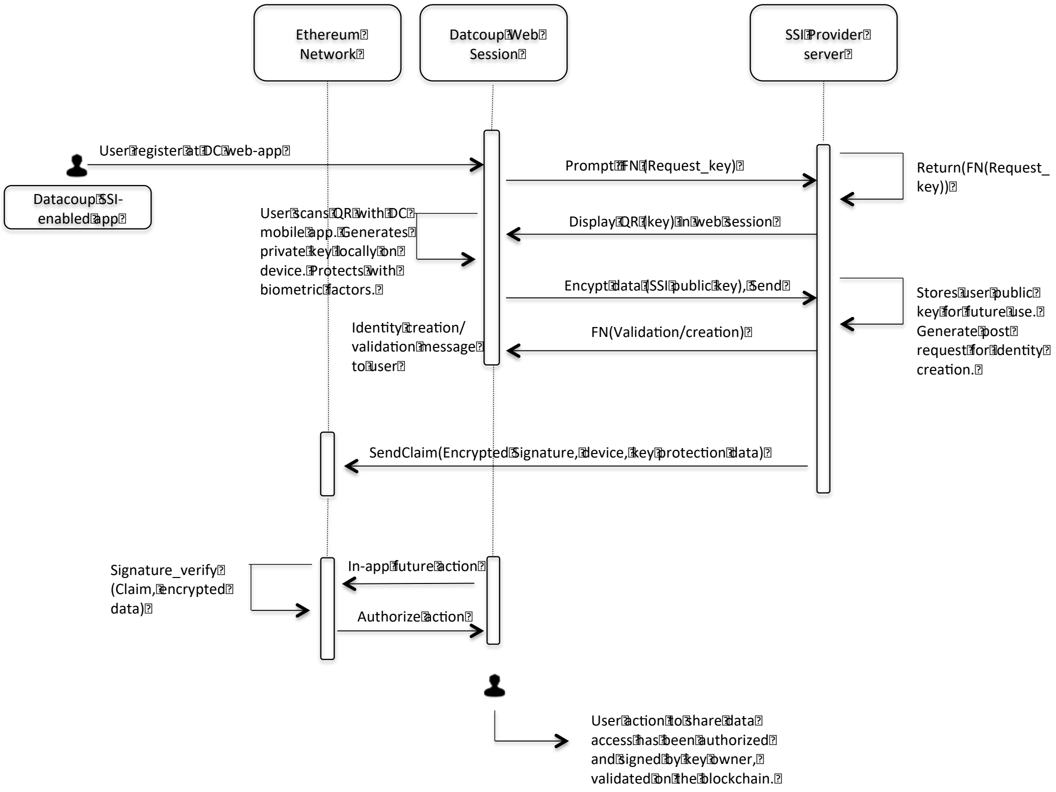
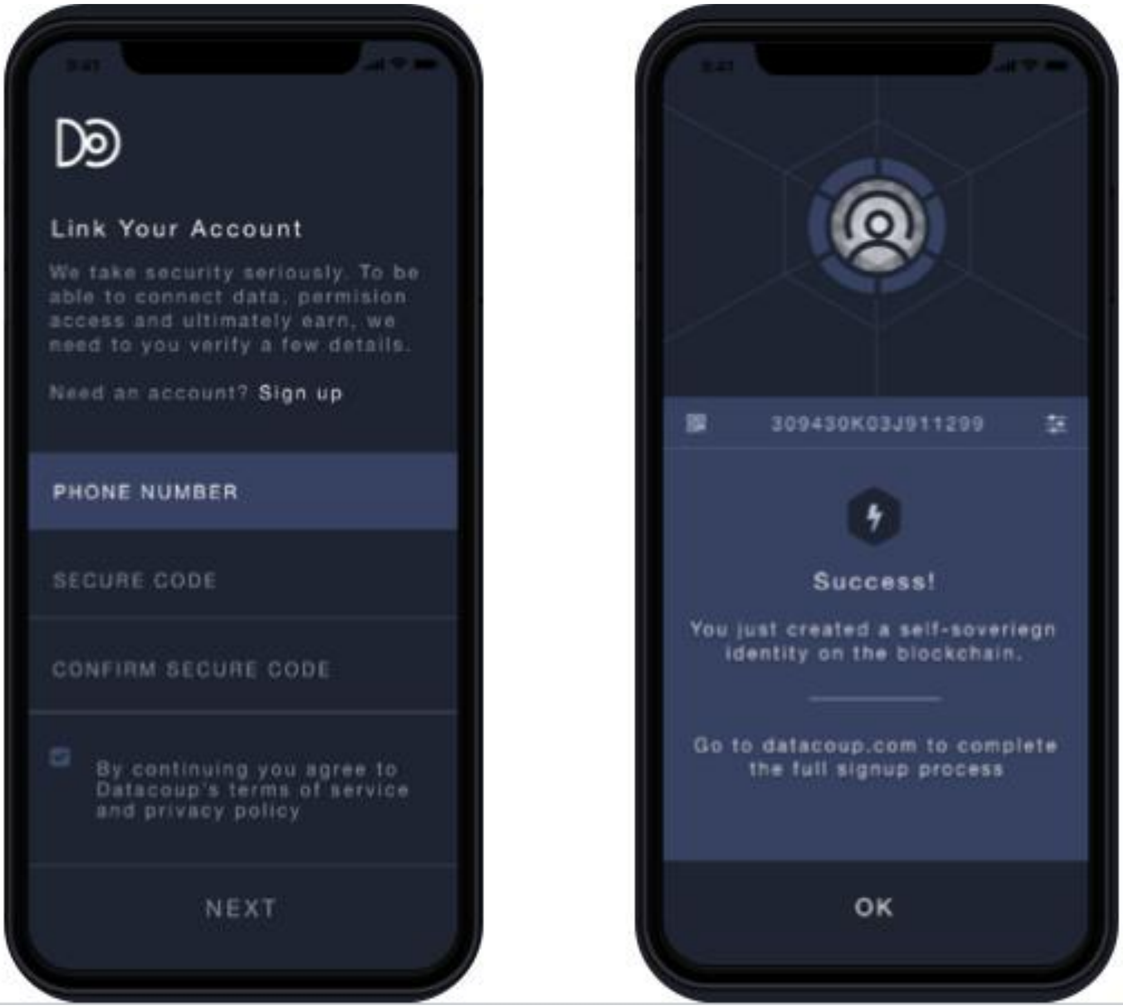

Ethereum Network
Datcoup Web Session
SSI Provider server
User register at DC web-app
Datacoup SSI-enabled app
Prompt FN (Request_key)
Return(FN(Request_key))
User scans QR with DC mobile app. Generates private key locally on device. Protects with biometric factors.
Display QR (key) in web session
Encypt data (SSI public key), Send
Stores user public key for future use. Generate post request for identity creation.
Identity creation/validation message to user
FN(Validation/creation)
SendClaim(Encrypted Signature, device, key protection data)
Signature_verify (Claim, encrypted data)
In-app future action
Authorize action
User action to share data access has been authorized and signed by key owner, validated on the blockchain.

Figure 5: Sequence Diagram - Identity

When a data provider begins registration with the Datacoup web-app, they are prompted to download the Datacoup mobile client. The provider then verifies that they own communication lines to the device via phone number and email multi-factor-authorization. With the Datacoup mobile app, a data provider scans the QR code shown on the web-app request_key. The QR itself is a unique public key, and when scanned, a private key generate_key is created on the provider's phone.



Link Your Account

We take security seriously. To be able to connect data, permision access and ultimately earn, we need to you verify a few details.

Need an account? **Sign up**

☑ By continuing you agree to Datacoup's terms of service and privacy policy

309430K03J911299

Success!

You just created a self-soveriegn identity on the blockchain.

Go to datacoup.com to complete the full signup process

Figure 6: Mobile App - Identity Creation Screenshots

A signature can authorize any future actions taken by the data provider, such as adding data sets, or entering transactions, with this key. Actions requiring data provider signatures will largely be controlled by data provider preference settings. The data provider secures the key by setting up biometric authorization and a PIN. The biometric data, PIN, phone number, and email, along with the public key, are all encrypted on the device, and stored with the key server for future use. Upon completion of these setup tasks, the web-app requests a post-back generate_pr success message for affirmative identity creation. The encrypted signature, device ID and device ownership data are all written SendClaim to the blockchain.

4.2 Component 2: Decentralized Data Storage

Once an identity has been created, the data provider begins the process of connecting their data within the web-app. We'll maintain focus on component 2 in this section, but it is worth noting that from component 3 (Compute and transact), we are using an Enigma [8] client (JS library) in the web-app to encrypt data locally on the device, before it is passed to any data store. This means that component 2's primary function is to store encrypted data in a decentralized network. As a result, we are exploring decentralized databases and file stores like bigchainDB, Sia, IPFS, Storj and more to find a solution that is consistent and scalable. Once in storage, the data is pulled into a secure hardware enclave for compute/transact (component 3, below), and returned to storage once the stipulations of a transaction smart contract have been met. Our data store has a couple simple goals:

1) Receive and store data
2) Make data available when requested.

Here is a sequence diagram exhibiting how component #2 accomplishes its goals.

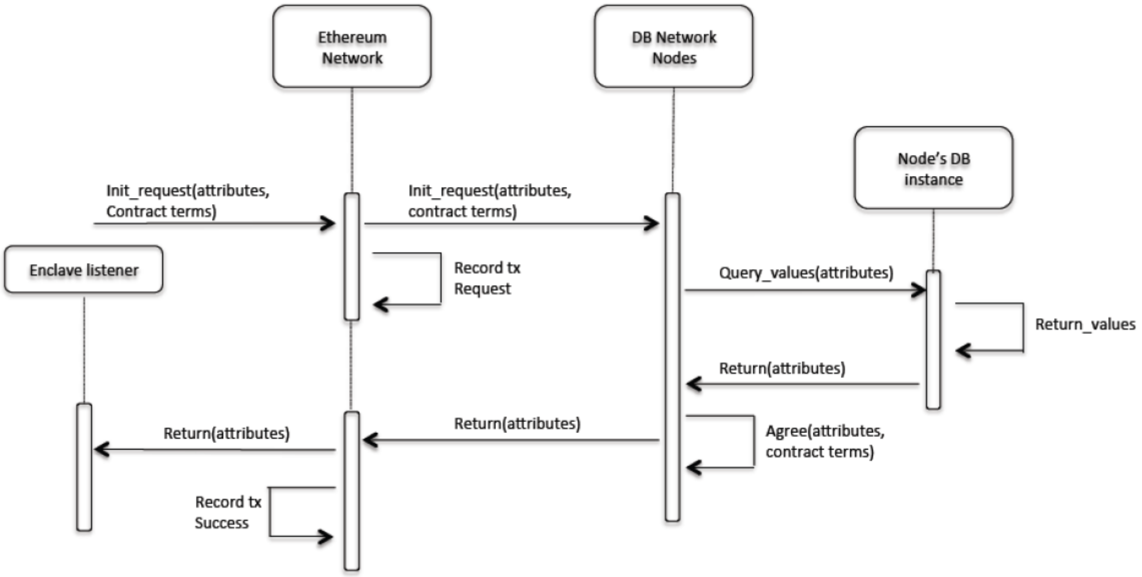


Figure 7: Sequence Diagram - Storage

As mentioned above, Enigma's (component 3) listener lives in the app and encrypts web-session data connections with the Enigma client public key. The data will be stored based upon a tagging ontology with categorization buckets. For example, we may push data into decentralized storage, tagged by demographic information, such as age, gender or location. These simple tags alleviate the need to pull all data into the secure enclave anytime a compute function is requested by a data requester, while also minimizing data leakage of potentially sensitive data. This significantly cuts down on the computation load, while maintaining sufficient privacy.

When a transaction smart contract has been mobilized, its new active state is logged on the blockchain. It then will initialize a request init_request(contract terms) with the database nodes, calling for data pertaining to the contract. Once the nodes agree that the request is valid, database instances are queried for the data query_value(attributes). Once values have been returned, nodes agree that the query meets the stipulations in the contract agree_terms, and log the successful return of attributes based upon the request. The values are sent to Component 3's secure enclave, for compute functions return_attributes, and the data is ready to be analyzed.

4.3 Component 3: Decentralized Computation

Traditionally, in order to use consumer personal data, data requesters such as hedge funds require access to the consumers' raw data (which they currently buy from 3rd party vendors). In contrast, our design will allow data requesters to utilize the data providers' raw data without gaining access to the decrypted raw data itself, thus preserving privacy. We achieve this "magical" property with our system's 3rd key component: the Enigma protocol for "secret computation". Enigma provides a decentralized network that allows computation in a TEE (Trusted Execution Environment) with strong correctness and privacy guarantees. The Enigma network is similar to Ethereum, but with the key difference that the decrypted data itself is concealed from chain and even from the nodes that execute the computations. Data requesters can purchase rights to compute over the data (analyze it) instead of buying the data itself.

For example: A hedge fund data requester could find a correlation between demographic profiles and Starbucks purchases, without gaining access to any single raw consumer data. Nor will any other party gain such access (including neither Datacoup nor nodes that execute the computation).

Here is a sequence diagram exhibiting how component #3 works:

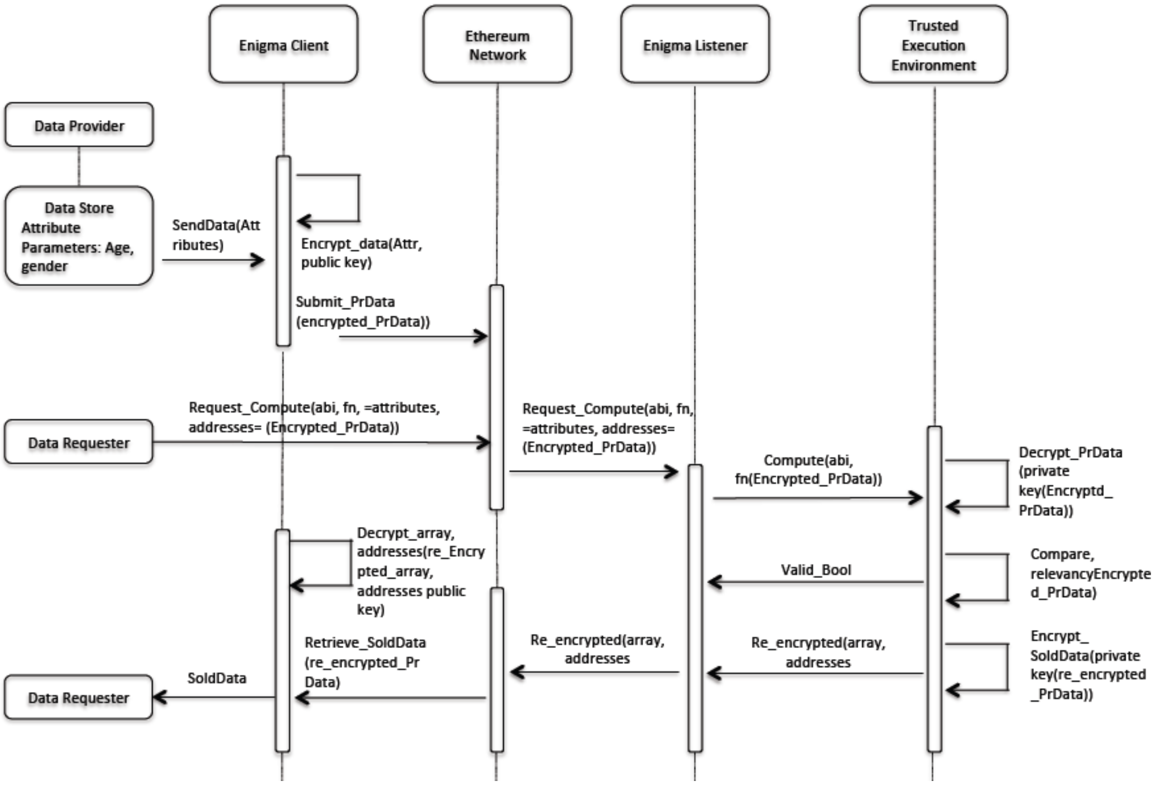

Enigma Client
Ethereum Network
Enigma Listener
Trusted Execution Environment
Data Provider
Data Store Attribute Parameters: Age, gender
SendData(Attributes)
Encrypt_data(Attr, public key)
Submit_PrData (encrypted_PrData))
Request_Compute(abi, fn, =attributes, addresses= (Encrypted_PrData))
Data Requester
Request_Compute(abi, fn, =attributes, addresses= (Encrypted_PrData))
Compute(abi, fn(Encrypted_PrData))
Decrypt_PrData (private key(Encryptd_PrData))
Decrypt_array, addresses(re_Encrypted_array, addresses public key)
Valid_Bool
Compare, relevancyEncrypted_PrData)
Retrieve_SoldData (re_encrypted_Pr Data)
Re_encrypted(array, addresses
Re_encrypted(array, addresses
Encrypt_ SoldData(private key(re_encrypted _PrData))
SoldData
Data Requester

Figure 8: Sequence Diagram - Computation

As mentioned in Section 3.2.2, the data provider private data is stored in a decentralized database only after the Enigma client (JS library) encrypts it. The data could only be decrypted inside an Enigma TEE (enclave). A data requester purchases the right to compute over the data by initiating a transaction in our marketplace {inputs: data fields required, sample size, qualifying data attributes, computation functions, gas limit, computation fee, payment for data, address for analysis result}, the smart/secret contract pulls the required encrypted data from the decentralized database into an Enigma network enclave. Inside the enclave, the data is decrypted, and the computation/analysis is executed. The computation result is encrypted in the enclave, and sent to the data requester.

4.4 Marketplace Smart Contract

If we zero in on the Ethereum vertical line in the above POC sequence diagram (Figure 8), here are the inputs and outputs from the smart contract:

Inputs:
Submit_PrData(encrypted_PrData)
Request_Compute(abi, fn, =attributes, addresses= (Encrypted_PrData))
Re_encrypted(array, addresses)
Outputs:
Request_Compute(abi, fn,=attributes, addresses=(Encrypted_PrData))
Retrieve_SoldData(re_encrypted_PrData)

When a data requester enacts a purchase transaction, they send the smart contract a computation request Request_Compute(fn, providers_data_fileds_required) along with payment. The data provider sends his encrypted data to the smart contract: Submit_PrData(encrypted_PrData). Then the smart contract sends out a computation request to the Enigma network Request_Compute(fn, encrypted_PrData). An Enigma worker inside a TEE does the computation and the computation result is sent back into the smart contract Compute_Result(re_encrypted_result). The smart contract then transfers the computation result back to the data requester Retrieve_SoldData(re_encrypted_result). Finally, the smart contract sends out the data requester's payment to the data provider.

5 Ecosystem Token

5.1 The Reasons for PDT

There are several reasons why the PDT ecosystem requires its own token:

- Isolating Demand for Consumer Personal Data:
With a native ecosystem token dedicated to personal data transactions, demand for the token represents demand for the underlying data. The ecosystem is, in effect, backed by the value of the data and future data transactions of the platform at scale. Using an existing token like Ethereum, the value of which is derived from its status as a smart contract platform, would obfuscate the true value of the underlying data resource. PDT serves to isolate and atomize the value of the tradable ecosystem resource: consumer personal data.

- Price Signaling and Transparency:
Consumer ownership and transacting of data is a nascent behavior. Current markets for data are extremely fractious, with businesses demanding data in different forms and "paying" for it in different ways:

- Facebook "pays" the consumer with a "free" app
- A life insurer "pays" by lowering policy premiums
- Amazon keeps consumer order histories, but obfuscates the costs of data collection in the prices they charge consumers for goods and shipping

- A data broker surreptitiously takes consumer data without paying

A native token unifies disparate and idiosyncratic demand for data access into a single homogenous pricing unit. Armed with a clear price signal, the consumer can make rational decisions with regards to selling data. This is critical in a nascent market where price discovery is already an existential market dynamic.

- Early Participant Incentives:
Data providers - PDT ownership serves is an access "key" to participate in the data marketplace. Data providers must earn and hold PDT in order to participate in ongoing data sales. So, for data providers, the earlier they connect and make data available to requesters, the more PDT they can accrue and hold for ongoing, long-term data sales and other benefits.

Data Requesters - Initially, demand for data should rise commensurately with the amount of data available in the marketplace. The earlier requesters purchase and hold PDT to access the data marketplace, the lower the cost of access.

- Transaction Cost Reduction:
The PDT ecosystem caters to data providers and requesters from all over the globe. It does not operate in a single region, or location. There are network participants from disperse locations, representing many different fiat currencies. A native token significantly reduces data transaction costs and friction across a global participant network.

- Transaction Auditability:
Building on the above point, blockchain-based records provide an immutable audit trail of data transactions. Token payments provide essential details like who bought data and when they bought it. Trails of token ownership bring necessary transparency and oversight ability to a nascent industry that has thus far been plagued by unaccountable market actors (i.e.; 3rd parties).

5.2 Token Economics

For any currency, overall demand is composed of use demand and savings demand. Both are fundamentally forms of savings that differ simply in horizon, analogous to dollars held in a wallet (use) vs. dollars held in a bank account (savings). Empirically, most demand for currencies, whether fiat, commodity, or digital, are in the forms of savings, so savings demand is ultimately of key importance to token price.

5.2.1 Usage Demand: Data Requesters

Initially, PDT use-demand will come primarily from data requesters, due to staking. In order to access the data markets, requesters must buy and hold tokens. Based on Datacoup's existing business, we see solid demand among:

- Investment managers - To use anonymized personal data to analyze investment hypotheses
- Market researchers and retailers - To guide market research and product development
- Marketing intermediaries - To test conversion rates and fine-tune campaigns.

We expect these current requesters of personal data to increase usage simply as a result of lower transaction-costs and superior data. And, of course, any regulatory action like GDPR [9] against non-permissioned data collection would grow this market further. Additionally, we anticipate larger economic verticals – insurance, credit and lending, telecom – to ultimately enter the market, which will increase demand even further.

5.2.2 Usage Demand: Data Providers

In order to participate in the data marketplace, data providers will be required to hold minimum requirements of PDT that they have received from connecting data. This will have the effect of minimizing liquid token supply, as providers are incentivized to hold tokens. Additionally, as more apps, products and services are introduced on the platform, the overall demand for monetization will manifest in more token lockups and increased demand for the token in order to monetize.

5.2.3 Usage Demand: All Participants

Because the PDT token is the only way to exchange on the PDT marketplace, its core terminal use-demand comes from data requesters buying data and from data providers and spending PDT on apps and consumptive products and services in the ecosystem. So long as requesters continue to wish to reach providers on the platform, those requesters will continue to offer fiat for PDT, giving the tokens sustained value. That sustained value, in turn, incentivizes providers to both accept and to save the tokens they receive, confident they can use them in the future.

5.2.4 Savings Demand

Hand-in-glove with usage demand, PDT savings demand will initially come from data providers and requesters maintaining balances. As third-party service providers join the platform, both sides of the market, and the data service providers themselves, will all tend to increase savings demand, ceteris paribus, in proportion to their use of the platform. Because savings are so powerful in protecting token price, the token is designed to encourage savings. As stated above in section 5.2.1, staking can be used to encourage savings, and is appropriate where the asset being bid on is commodity. As such, staking is available to three groups on the PDT platform: data providers, data requesters, and third parties.

5.2.5 Savings Demand: Data Provider Minimum Requirements

In the case of data providers, the ecosystem will require minimum balances in order to participate in the marketplace. This is primarily to encourage maintenance of data connections and is incentivized by the potential for future earnings.

5.2.6 Savings Demand: Data Requester Staking

In the case of requesters, the ecosystem will require token purchases and holdings in order to access the marketplace. Additionally, premium access to specific data or data providers can be used to incentivize greater amounts of PDT holdings.

5.3 Token Supply

Unlike a security, PDT tokens do not offer ownership of an asset, rather access to a platform. Therefore the supply component of token price that is most relevant is the number of liquid tokens. In order to directly impact price, a given token must have 3 characteristics, each describing a smaller universe of supply: the token must be issued (not planned); it must be liquid (not locked), and it must not be saved by the owner (not saved) [10]. Because savings will be dynamic and unpredictable, in practice the best aggregate we can predict is liquid token supply. Please see the PDT Token Addendum for a closer look at liquid supply projections.

5.4 Ecosystem Token Flows

As outlined in section 5.1, token flows are driven by holdings and usage of PDT. The below diagram (Figure 9) outlines these basic notions in the form of an Exchange Layer (usage demand) and a Staking/Withholding Layer (savings demand):

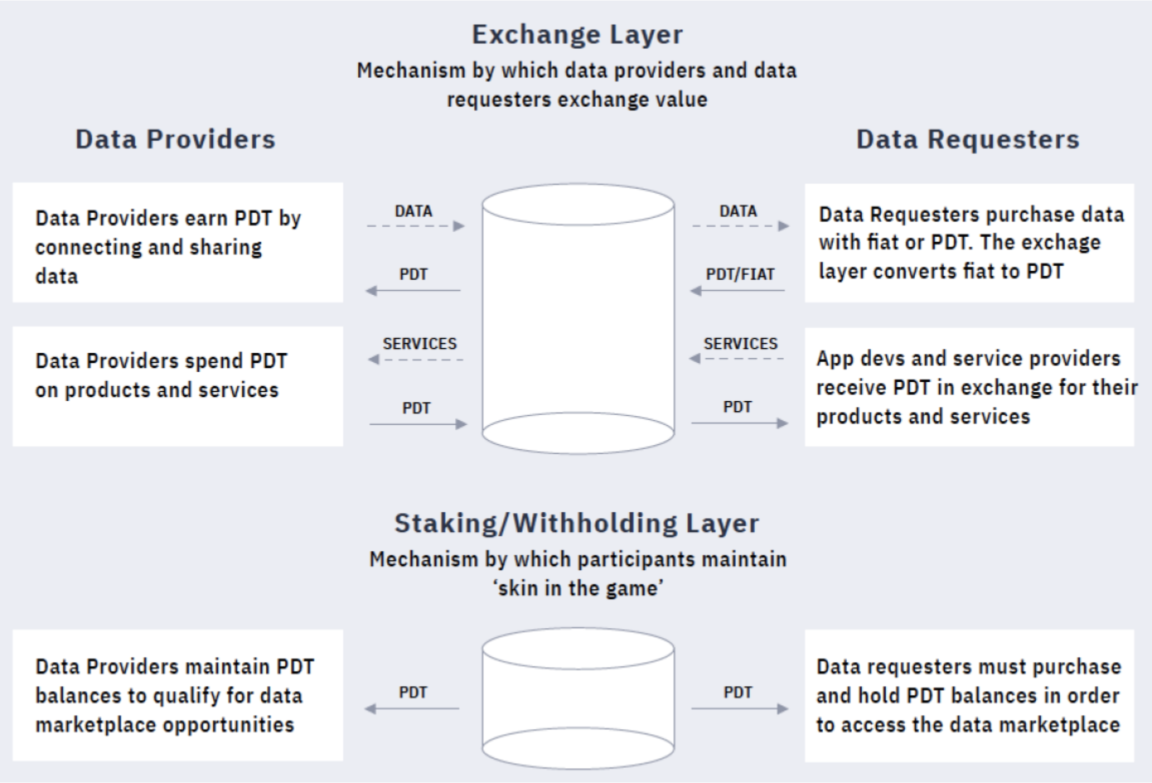

Exchange Layer
Mechanism by which data providers and data requesters exchange value
Data Providers
Data Requesters
Data Providers earn PDT by connecting and sharing data
DATA
PDT
DATA
PDT/FIAT
Data Requesters purchase data with fiat or PDT. The exchage layer converts fiat to PDT
Data Providers spend PDT on products and services
SERVICES
PDT
SERVICES
PDT
App devs and service providers receive PDT in exchange for their products and services
Staking/Withholding Layer
Mechanism by which participants maintain 'skin in the game'
Data Providers maintain PDT balances to qualify for data marketplace opportunities
PDT
PDT
Data requesters must purchase and hold PDT balances in order to access the data marketplace

Figure 9: Tokenomics - Exchange and Staking Layers

6 Conclusion

Our proposed ecosystem will provide consumers with increased control over and ownership of their valuable personal data. It will also alleviate many of the ills that currently plague the current state of personal data on the centralized web.

The PDT ecosystem is designed to isolate the underlying resource of personal data and transfer its current centralized-world value into our decentralized economy. The PDT token serves as an "access key" to the ecosystem and a medium of exchange for data transactions.

Empowering consumers to own and transact with their data is a natural extension of the overall decentralization movement that started with Bitcoin in 2009. Our aim is to break down the walls of central data silos and shift data control and the economic value of data back to the consumer.

References

[1] John Deighton, Peter Johnson. "The Value of Data 2015: Consequences for Insight, Innovation & Efficiency in the U.S. Economy". Commissioned by: *Direct Marketing Association, 2015*. URL: https://thedma.org/advocacy/data-driven-marketing-institute/value-of-data/.

[2] FTC (Federal Trade Commission). "The Equifax data breach". In: *FTC official website*. URL: https://www.ftc.gov/equifax-data-breach.

[3] Investis. "Equifax 2017 10K". In: Official SEC 10K filing. URL: https://otp.tools.investis.com/clients/us/equifax/SEC/sec-show.aspx?Type=page&FilingId=12595195-162329-176028&CIK=0000033185&Index=12200.

[4] Suzanne Barlyn, "Strap on the Fitbit: John Hancock to sell only interactive life insurance" (Sept 2018). URL: https://www.reuters.com/article/us-usa-wireless-fcc/u-s-regulator-demands-companies-take-action-to-halt-robocalls-idUSKCN1NA2KH

[5] Satoshi Nakamoto. "Bitcoin: A Peer-to-Peer Electronic Cash System". https://bitcoin.org/bitcoin.pdf, 2008.

[6] Connor Blenkinsop. "Blockchain's scaling problem, explained. In: *Cointelegraph* (Sept 2018). URL: https://cointelegraph.com/explained/blockchains-scaling-problem-explained

[7] NEO Team. NEO White Paper. URL: http://docs.neo.org/en-us/whitepaper.html.

[8] Enigma Team. Enigma Github Repository. URL: https://enigma.co/

[9] European Data Protection Supervisor, European Co-Legislators. *The European General Data Protection Regulation* (May 2018). URL: https://eugdpr.org/.

[10] Murray Rothbard. "The pattern of indirect exchange", Chapter 3. In: *Man Economy and State*. David Van Nostrand, 1962, pgs 187-213. ISBN: 978-1-933550-27-5