“What is needed is (a)... system based on cryptographic proof instead of trust, allowing any two willing parties to transact directly with each other without the need for a trusted third party.”

-S. Nakamoto
“Bitcoin: A Peer-to-Peer Electronic Cash System”

1. Introduction

Blockchain was built to solve the issue of trust in an online environment. An Economist article went so far as to dub blockchain, the underlying technology to bitcoin, the "trust machine." The key innovation of blockchain is the ability to sign, validate and immutably record a transaction without a third party. In the financial system, third parties can be costly, if often necessary, intermediaries of the flow of financial information and value. Personal information and “in person” proof of identity is traditionally needed in order to establish this trust.

Blockchain technology, on the other hand, was designed as a way to manage trust via a cryptographic protocol run on a peer-to-peer network. The protocol and network (enabled by blockchain technology) protect buyers and sellers by creating an immutable version of reality as it happened without the need for a trusted third party, such as a bank or escrow services. This new form of trust can be established an online service without “bricks and mortar” trust. The blockchain prevents fraud by creating an immutable, unchanging and trusted record record of what happened and when, agreed upon by a peer-to-peer network.

Similarly, our current healthcare system and the healthIT infrastructure to support it are, by historical necessity, “bricks and mortar” and institution-centric. These systems were largely designed and implemented in a pre-internet era to support the bricks and mortar of hospitals and clinics. These systems largely contain information collected around in-person visits. As we enter an era of personalized medicine, we need systems designed to track a person’s health, to understand disease with each individual’s unique context in the course of both disease and wellness. We must have systems that can track an individual’s data longitudinally throughout their lives independent of location. To have a truly patient-centric culture that will drive precision medicine and patient centered outcomes research (PCOR), we need a truly patient-centric healthIT infrastructure built on trust.

Trust is a critical, if underappreciated, component of information flow in healthcare. It is often unclear who among third parties has the rights at any given time to share and exchange health data. So default becomes a lack of sharing, no matter what the current technological limitations of sharing.

Dr. David Kibbe, CEO of DirectTrust recently articulated the problem via HealthcareITNews, “We’ve got the wrong problem in mind, and it’s distracting us from getting to the root causes of our discomfort...The real issue involves who owns the health information that is created by you

and me, and about you and me, and who has the rights to view it, access it, use it, download it and move it around. The answer to the problem lies in returning the power to control medical records to the patients...” Another way to put it: the problem may be in the third party-driven architecture.

Last fall, in Nature Biotechnology, Kish and Topol argue that we we need patient control of health data and a trusted place to store and manage it. They argue blockchain technologies might provide some of the core components for a patient-centered system. Kish and Topol summarize our current state, “We’re looking at the prospect of a new, high-definition picture of individual human beings, and at the same time for that person’s data to be homeless, dispersed and inaccessible. Where the data live will determine the maps we can create and the directions we can go in with health, both as individuals and as a society.” In economics, clear individual ownership of assets allows capital to flow more freely. Similarly, Kish and Topol argue that with clear rights and individual control of health data, data will flow far more freely and securely that it does today. The alternative future scenario, whereby others control our data and know more about us than we can know about ourselves, is untenable.

To enable precision medicine, we need a place to store and individual’s health information that is free of third party control. Such a system way for individuals to store health data anonymously (not tied to personal identity information) such that it can be exchanged in a trusted fashion for medical research. Blockchain-related technology supplies the key ingredients for such a system. While current blockchains are slow and cumbersome for traditional data storage, we propose the same elements and benefits in a system designed for responsive and secure health data storage and exchange.

2. Proposed Solution

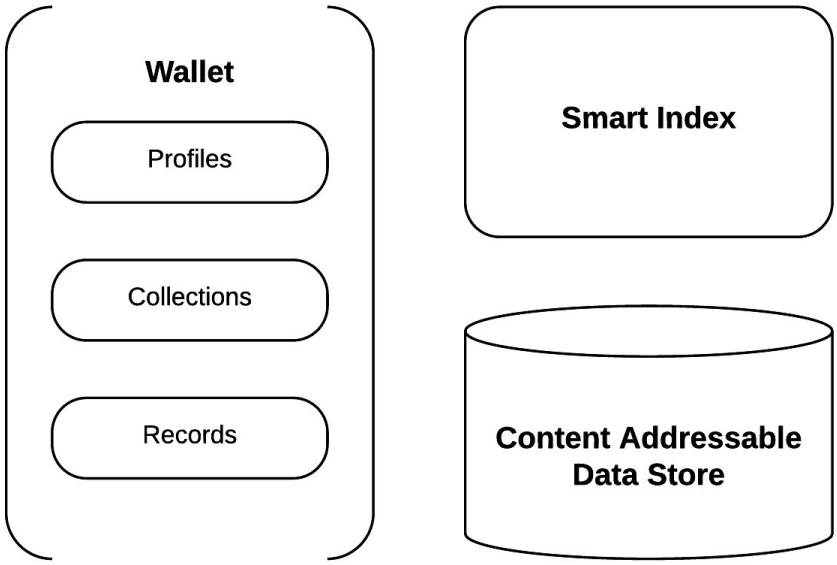
YouBase is a trusted, privacy-enabled, peer-to-peer data layer for health information storage and exchange built on blockchain technology. Using encryption, digital signatures, digital wallets, and distributed data stores, we provide the framework for management of personal and medical data that’s individual-centric. The core technology allows each personal data element of to be assigned a blockchain-compatible personal wallet address that cannot be linked back to an individual without the proper private key.

This framework provides several benefits, including: 1. a way to securely input, access and share any kind of file or record 2. a way to organize authorized access to information into a structured hierarchy 3. improved anonymous information sharing that could be used as a public or shared private asset 4. using this addressing, information sharing transactions can be tied to financial transactions.

With these tools in place, we imagine a future scenario where, rather than storing personal data, third parties could simply subscribe to data owned and controlled by the individual.

3. Implementation

YouBase introduces a new type of hierarchical, deterministic (HD) bitcoin wallet which uses BIP32 public-private key pair trees as addresses for an individual's structured data. We couple this wallet with persistent peer-to-peer content-addressable key-value stores, such as IPFS. The critical elements are the wallet, a smart index and a content-addressable data store.



The new HD wallet (client software) implementation is responsible for maintaining the BIP32 nodes defining the hierarchical structure, metadata, and pointers to the actual data, as well as managing permissions and handling encryption/decryption.

We propose a distributed hash table for holding this BIP32 tree metadata, for easier access, backup, and sync. The peer-to-peer key-value store is treated as just that, a "dumb" key-value store, so technically it could be hosted locally, on the cloud, or on a peer-to-peer file system such as IPFS.

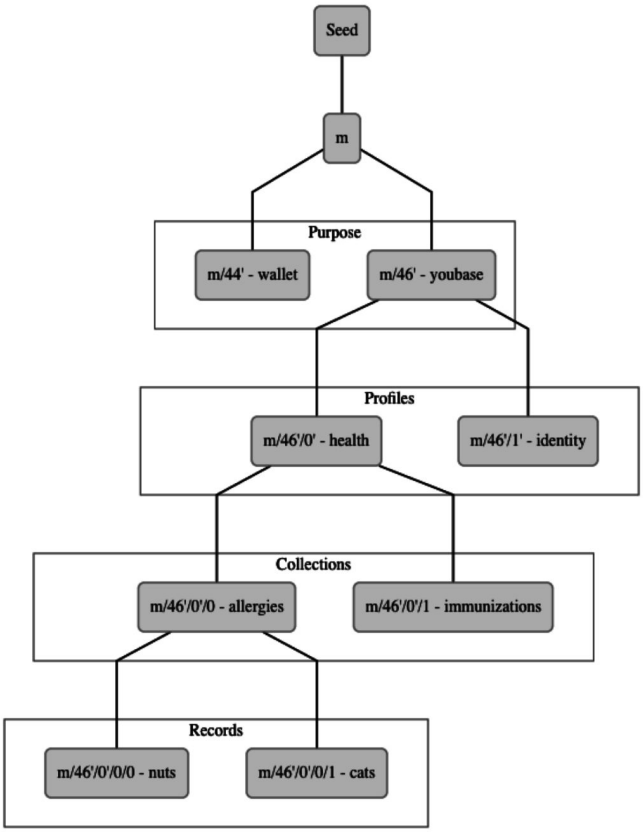
Structured Data

Our HD Wallet contains a tree structure with extended keys such that each parent key can derive the children keys, children keys can derive the grandchildren keys, etc. An extended key consists of a private or public key and a chain code. Sharing an extended key gives (private or public) access to the entire branch. A useful application is that a user can provide an extended private key to a trusted source that can then write (deposit) information in that branch of the tree without having access to information in other branches.

Providing public/private key pairings for structured data offers a number of benefits. First, specific branches can be used as data stores for specific types of information. Information secured in wallets can follow a pre-specified structure where different kinds of information are stored in different branches of the tree.

Second, access rights can be managed such that different parties can have different read/write permissions to single nodes or to entire branches. Data can be partitioned into separate branches, allowing users to grant access on a granular level, down to the individual record, or even changes to a record. Specifically, a private key has read/write access to the data while a public key has read access only.

Third, HD Wallets are flexible in that they can create sequences of public keys without having access to the private keys, so that read-only or receive-only permission can be granted in less secure environments without risking access to the private keys. From the outside (with access to a public key), there is no indication that the key is part of any larger structure. Each node becomes a bitcoin address like any other and seemingly unrelated to any other.



Finally, using bitcoin wallet structure means data exchange is enabled by allowing transfer just about any type of data from one party to another via JSON along with the ability to tie payments to information exchange by tying a hash of the data exchange to a bitcoin transaction. Think of a universal, secure email for data with unique addresses that are verifiable, but private, must be digitally signed, and can include a payment or used for data provenance, a record of rights and access to data at specific points in time.

Creating a Wallet

HD wallets are created from a root seed, meaning they can be backed up, restored, exported and imported by transferring the root seed. Seeds can be represented through a mnemonic sequence so they are easy to transcribe and remember via BIP39. The root seed acts as the input into a hash algorithm to create the master private key, chain code and master public key. The client software will manage the HD Wallet, including the encrypted seed and top-level master key pair. This client wallet software could exist on the web, smartphone or local computer. Depending on where the client software resides, wallet access can be granted via a number of methods to create the seed, including: photos, biometrics and multiple other means.

3a. Implementation: Structure of the YouBase Wallet

Profiles

The profile level of the wallet is also hardened and is used to define the structure of the nodes under it. Profile nodes point to a profile definition which the client uses to build the interface and interpret meaning of collections and records.

A YouBase wallet can have multiple profiles of the same type, such as health profiles. A person may have multiple health profiles for example, one for each provider. This allows the user to grant different levels of access depending on the context the user defines. Since each profile is hardened and on an isolated branch of the tree there is no way to know they are related without the owner of the profile's permission. Although stored and accessed completely separately, the client software may group together, or combine, related profiles. Since new profile types can be easily created, the collections are not limited to any one standard. A profile could exist for each data standard, say for HL7, BlueButton, FHIR and more. Profiles can also be created “on the fly” with JSON schema definitions.

Collections

Collections are specified in the profile definition of the collection's parent node (profile) and are used to hold groups of similar records. In the previous diagram we used a health profile which defines an allergy *collection* and an immunization *collection* as a simple example.

Records

The record is the base unit of storage in YouBase. The format of these records is also specified in the profile definition and is defined using JSON Schema, a standard for describing easily validated JSON data formats. Using this standard, profiles are not restricted to using a single format, providing the flexibility necessary to implement any conceivable application.

Schema

YouBase's JSON schema enables the creation of a profile definition that matches any existing format, easing the integration process while also helping clients understand the data by presenting it in a familiar way. We see interoperability then could be enabled on an individual

level rather than an institutional level. Over the long term, we believe this will speed health data interoperability. Individuals will have the incentive to drive interoperability of their own data.

The profile definition is the blueprint for a profile. It defines what collections a profile contains, how to validate a collection's records, what to encrypt, and how to interact with the record. At the top level, the profile definition only includes a title, type, and an array of collections. Collectionssd could be defined as FHIR Resource Type definitions, for example.

In our general schema, the type is used to group different types of related profile definitions together.

Example:

```
{
"title": "Health Profile",
"type": "health",
"collections": [...]
}
```


A profile's collections are defined as an array of collection objects, each having a title, type, schema, and form. A collection object's position in the array determines what branch it maps to under a profile. In a collection the type field is used to describe the collection's records in a consistent manner. This allows competing standards to create profiles that have a shared vocabulary.

Example:

```
{
"title": "Health Profile",
"type": "health",
"collections": [{
"title": "Allergies",
"type": "allergy",
"schema": {...},
"form": [...]
}, {
"title": "Immunizations",
"type": "immunization",
"schema": {...},
"form": [...]
}]
}
```


The schema field follows the JSON Schema standard and is used to validate a record. It includes a title, JSON Schema type, properties, and list of required fields. The title is the

singular name for a record as opposed to the plural name defined in the collection title. Unlike profiles and collections, the schema type must be defined as 'object' since we expect every record to be an object. This ensures the schema works with JSON Schema validators.

3b. Implementation: Distributed Storage

Using an HD Wallet to hold the structure of the data provides flexibility and allows a user to keep YouBase documents in a local filesystem, cloud storage, or our proposed distributed hash table. At the same time, it allows a user to easily sync documents between different storage systems.

Generically, YouBase describes the storage of two types of data: public-key-addressable, and content-addressable. BIP32 nodes are stored as signed JSON documents, where each node is addressable by its public key hash, or what would be a bitcoin address if nodes used to transact on bitcoin's (or another's public or private) blockchain. These documents contain links to external content which, in contrast to the BIP32 nodes, is addressable by the hash of the content. Data can be stored at every node in the HD Wallet tree by creating a signed JSON Document. The public and private keys are not stored in the document but are used to secure the document and control access to it. Large amounts of data, say medical images, can be stored at the node by using link storage.

The document is stored in a DHT where the lookup key is an address generated from the public key. The document is saved as JSON and contains the address as an id, revision, document data, document links, and a signature.

Example:

{
"_id": "address generated from public key",
"_rev": "hash of data + links + _lastrev",
"_lastrev": "points to previous _rev if this document is an update",
"data": {...},
"links": [...],
"notaries": {...},
"signature": "signature of _rev"
}

The document's id is an address generated from the public key and is a valid bitcoin address. It is used as the lookup key but also included in the saved document to make syncing and validating easier.

Data

The data field is the core of a document, containing: timestamps, references to profile definition, issuer, signatures, demographic information, and the actual information to be stored. The data

field follows a standard format defined in a specification document. The data field can also reference links.

Example:
{
"_id": "address generated from public key",
"_rev": "hash of data + links + _lastrev",
"_lastrev": "points to previous _rev if this document is an update",
"data": {
"type": "record",
"profile": "hash of profile definition",
"issuer": "address of entity creating the record",
"timestamp": "2015-05-21T20:21:46.612Z",
"demographics": {
"geohash": "9xj",
"gender": "female",
"age": 24
},
"record": {...} // record data as defined in the profile definition
},
"links": [...],
"notaries": {...},
"signature": "signature

Signature
To ensure a document is valid it must include a signature of the document hash by the node's private key. When a document is saved the document store simply checks the signature against the document hash and id.

Link Storage
The link storage is a key value store where the lookup key is a hash of the contents. This can be as simple as a file saved to the local file system, where the file name is a hash of the file content or a proprietary cloud solutions like Amazon S3 where, again, the file name is a hash of the contents. IPFS is the logical choice for YouBase at this time as IPFS fulfills current requirements for a distributed store, independent of any one 3rd party.

4. Applications and Use Cases

Security: Security breaches in healthcare have become all-too-common, over 100 million health records were breached in the US in 2015 alone, because of the high value to identity thieves. Providing a repository of data independent of a third party will provide a framework where personal data is provided on a subscription basis, rather than stored in multiple locations

by multiple third parties. While no security solution will be perfect, limiting the size of large “honeypots” of personal information has the ability to limit exposure to large-scale breeches.

Access. At present, a person's ability to access and manage personal information that resides with third parties is challenging. Part of the problem is that no single repository that can accept health data for an individual. With our proposed framework, each person will have a private, universal and secure container of their own medical data that serves as a reference for health care stakeholders. Data can be stored as a document, largely independent of the format.

Interoperability. YouBase is data agnostic and could act as a single repository of data independent of the data types stored in each profile. A data profile can be accepted and defined in the schema, so there's no need to pre-define the data structure before receiving the payload. In this way, interoperability could be a more open marketplace, whereby developers can create solutions to make data interoperable between profiles as demanded by individual users.

Population Health and Biomedical Research. As each data element is not directly tied to personal information, individuals can anonymously donate their validated data with minimal metadata or personal information.

Anonymous messaging in research: If some anonymous data is found to be valuable, say, a unique biomarker found during the course of clinical research, researchers could message the data owner without being aware of the data holder’s identity. Because each data element has a unique address, YouBase facilitates the ability for researchers to send a message to a data element, maintaining anonymity, but allowing for communication.

Sending records across systems. A set of universal addresses for data elements will allow for secure transmission of a patient record or other, simply by scanning a public address for which the health care provider has a private key.

A universal ID system with Identification at point of care. YouBase can provide validated identity at the point of care through our wallet IDs shown in a QR code on a phone. For example, a user could enter a lab and show that his identity matches with a traditional ID, or simply present a YouBase token that has been signed. The public key token is then verified as belonging to the user and documented in the system.

The phlebotomist takes a blood sample, signs the date/time, and the sample is then permanently associated with the user's token/private key. The sample is processed using the lab's existing system. The results (data) are sent to the user's YouBase wallet and can only be opened/viewed with the token/secure-key.

This could be applied to many health care transactions, creating validated identity and a universal set of addresses through which information could be exchanged. Using this kind of

system will also improve data quality as each data entry will require a digital signature linked to the person who entered the information.

5. Feasibility

We have shown feasibility with a working proof-of-concept and are currently bringing the system to scale. Our first project will launch in September, 2014 with a genomics data exchange and aggregation. A leading research center will use YouBase's technology to build a world class data sharing platform, focused on cardiovascular chronic illness. The initial project, focused on heart transplant patients, is expected to go live Q1 of 2017.

6. Summary

Using an HD Wallet to access a personal data store which is provides:

- **A Universal repository with granular access controls.** Securely input, access and share any kind of file or record by providing keys and digital signatures. A way to organize and authorize access to information in a structured hierarchy.
- **Privacy and sharing:** Authenticated data and proven identity without storing third party personal information. Without a high-level key, no one element of data can be tied to any other. With this level of data anonymity,more data can be donated by users for use in various forms of data commons.
- **Blockchain available, but not dependent**. Information sharing transactions can be tied directly to a universal ledger as a public proof of provence and/or as a financial transaction, but are not required to be. Wallets can be used solely as structured access to a data store, and are also valid bitcoin addresses.
- **A set of universal unique IDs.** Using a public-private key hierarchy for data access provides for a universal address to read or write information based on an individual's HD Wallet.
- **Longitudinal tracking.** All transactions are time-stamped, so all records can become a longitudinal record. The universality of JSON means it's easy to accept data from multiple sources.
- **Security.** Encryption of data by default and because each data element has a unique address, any one breach can be insulated from an attack on another.

With YouBase, information and rights to one's personal information will *follow an individual* as she moves through various contexts in her daily life. A complete, longitudinal record can be created, including consumer-generated application data, with the individual as the primary controller of access - all independent of a third party. With this framework, YouBase seeks to enable PCOR, precision medicine and the interoperability roadmap, which states, "With the availability of….these measures consistently available throughout the health system, new interventions can be studied with minimal investment and comparative effectiveness information will be available for all therapeutic options. "

References

Nakamoto, S., “Bitcoin: A peer to peer eclectronic cash system”, https://bitcoin.org/bitcoin.pdf 2008

Editors, “The Trust Machine” , The Economist, http://www.economist.com/news/leaders/21677198-technology-behind-bitcoin-could-transform-how-economy-works-trust-machine, 2015

Manos, D.,“CIOs tell ONC: It's time to let the EHR market innovate to drive interoperability” http://m.healthcareitnews.com/news/cios-tell-onc-its-time-let-ehr-market-innovate-drive-interoperability, Health IT News, 2016

Kish, L.J. and Topol, E.J., “Unpatients—why patients should own their medical data” *Nature Biotechnology* **33**, 921–924 (2015)

ONC Nationwide Interoperability Roadmap, https://www.healthit.gov/sites/default/files/hie-interoperability/nationwide-interoperability-roadmap-final-version-1.0.pdf, 2015

Supporting Technologies:

BIP32 (specification)
BIP39 (specification)
BIP43 (specification)
BIP44 (specification)
BIP45 (specification)