HealthDex

WHITE PAPER

Version 1.2

Abstract

HealthDex is a first-of-its-kind health data blockchain platform and decentralised exchange. The HealthDex blockchain platform is the perfect choice for new health data decentralised applications and traditional health data companies wanting to migrate to the blockchain. The HealthDex decentralised exchange is an off-chain marketplace where health data vendors, consumers and other participants can easily search, connect, trade and utilise health data.

Authors

Fahad Khan
Dr. Mohsin Chaudhry
Frederik Bussler
Vishnu Devarajan
David Chen

Table of Contents

1 DISCLAIMER....4
2 EXECUTIVE SUMMARY 4
3 BACKGROUND....5
3.1 HEALTH DATA 5
3.1.1 Volume of Health Data....7
3.1.2 Types of Health Data 7
3.1.3 Health Data Considerations 12
3.1.4 Health Data Market.... 15
3.2 BLOCKCHAIN AND DISTRIBUTED LEDGER TECHNOLOGIES 16
3.2.1 Blockchain Health Data Companies 17
4 PROBLEM ANALYSIS 18
4.1 The Platform Problem 18
4.2 The Marketplace Problem 19
5 HEALTHDEX: A COMPLETE SOLUTION....20
5.1 STAKEHOLDERS.... 20
5.1.1 Data Vendors 20
5.1.2 Data Curators.... 21
5.1.3 Data Consumers 21
5.1.4 Other Marketplace Participants 22
5.1.5 Nodes 22
5.2 HEALTHDEX TOKENS 22
5.2.1 Node Remuneration.... 23
5.2.2 Speed and Micropayments.... 23
5.2.3 Market Proxy 23
5.2.4 Other Benefits 24

5.3 TECHNICAL COMPONENTS 24

5.3.1 Front End 24

5.3.2 Entity Resolution 25

5.3.3 Decentralised Key Management System and Encryption........ 25

5.3.4 Fully Homomorphic Encryption (FHE)........ 27

5.3.5 Federated Learning 28

5.3.6 Application Containerisation 31

5.3.7 Blockchain Layer........ 31

5.3.8 Data Layer........ 32

5.4 OTHER FEATURES 35

5.4.1 Entity Matching Algorithms and Blended Data 35

5.4.2 Data Management and Ownership........ 35

5.4.3 Data Security and Encryption 36

5.4.4 Fair Distribution........ 37

5.4.5 Data as a Transferable Asset........ 37

5.4.6 Fraud Prevention........ 38

6 REVENUE MODELS........38

7 REFERENCES........40

1 DISCLAIMER

PLEASE READ THIS DISCLAIMER SECTION CAREFULLY.

CONSULT LEGAL AND FINANCIAL EXPERTS FOR FURTHER GUIDANCE.

The following information may be incomplete and in no way implies a contractual relationship. While we make every effort to ensure that all information in this White Paper is accurate and up to date, such material in no way constitutes professional advice. HealthDex Limited neither guarantees nor accepts responsibility for the accuracy, reliability, current status (as of this White Paper), or completeness of this content. Individuals intending to invest in the platform should seek independent professional advice prior to acting on any of the information contained in this paper.

2 EXECUTIVE SUMMARY

The amount of health data in the world is growing rapidly and is expected to double every 73 days by the year 2020. [1] The multi-trillion-dollar health sector is critically reliant on health data for informed decisions. Health data are used to make new drugs and therapies, improve healthcare delivery, inform insurance and government policies, and much more.

Thousands of companies and organisations around the world buy and sell health data for billions of dollars annually. These individually-brokered deals are slow and expensive, meaning that huge amounts of data are underutilised. Currently, no health data marketplace exists where all stakeholders can easily search, connect, trade and utilise health data. Additionally, a new generation of blockchain health data companies are providing health data owners with greater control, security and remuneration of their data. Hundreds more of these health data decentralised apps (DApps) are expected to emerge but lack a blockchain platform uniquely tailored to the demands of health data.

HealthDex is a first-of-its-kind health data blockchain platform and decentralised exchange. The HealthDex blockchain platform is the perfect choice for new health data DApps and traditional health data companies wanting to migrate to the blockchain. HealthDex provides DApps with a range of on-demand enterprise solutions, such as decentralised data storage and security using Interplanetary File System (IPFS) cloud storage, as well as data processing using the uniquely tailored health database, HealthDexDB.

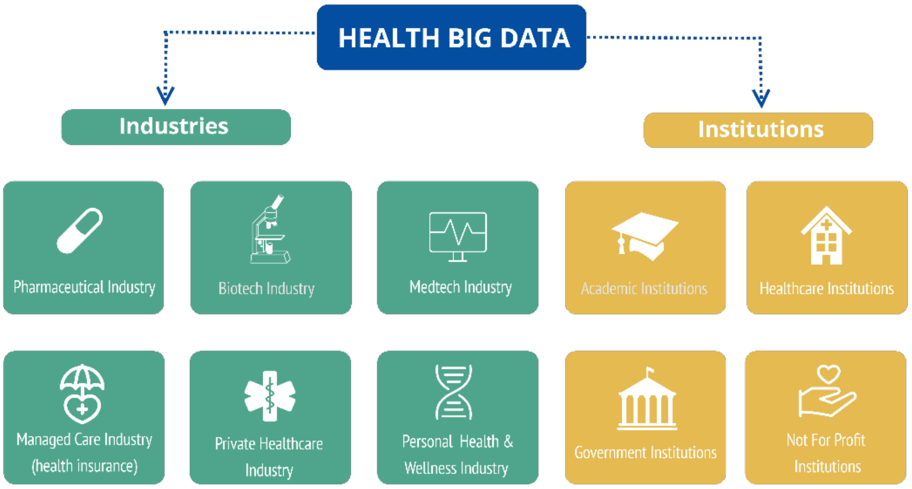
The HealthDex decentralised exchange is an off-chain decentralized marketplace where health data vendors, consumers and other participants can easily search, connect, trade and utilise health data. The HealthDex tokenised service layer enables frictionless transactions and remuneration, whilst ensuring appropriate incentivisation of nodes participating in the HealthDex network. Fully homomorphic encryption, federated learning and application containerisation technologies are incorporated for data handling and analysis. These allow data consumers to gain insights and train artificial intelligence models whilst minimising the need for data transfer and, thereby, ensuring sustainable and ongoing monetisation of health data. Because revenues are primarily generated from commission fees, HealthDex is incentivised to maximise the number of market participants and trading volume on the HealthDex platform. In the process, HealthDex will enable platform participants to generate ongoing medical discoveries, health insights and improved models of healthcare.

3 BACKGROUND

3.1 HEALTH DATA

The health sector is undergoing a revolution fuelled by ‘big data’. Whilst the health sector has always generated huge amounts of data, this data has only relatively recently started becoming available in digital form. This has triggered an avalanche of useful data that has traditionally been poorly managed. In the health sector and elsewhere, these massive quantities of data—colloquially called ‘big data’—are characterised by high velocity, complexity and variability. The growth of big data has rapidly outpaced storage, distribution, processing and analysis capabilities, while simultaneously kicking off an information revolution.

The health sector itself is a multi-trillion-dollar machine comprised of numerous industries and institutions, such as life sciences companies, academia, and government organisations. The availability of health data has shown great promise in tackling some of the greatest problems facing mankind today. At the same time, health data are overwhelming, not only because of the volume and diversity of data types but also because of the speed at which they are created. In response, new technologies, including blockchain, distributed ledger, cloud computing and data science techniques, are being increasingly utilised by companies and organisations to unleash the full potential of health data like never before.



Pharmaceutical, biomedical and other life sciences companies are using health data to discover the next generation of drugs, therapies, diagnostics and other medical products. Academics and clinicians are using health data to create new models of healthcare and to enable the next generation of healthcare delivery, termed ‘precision medicine’. Health insurance companies are using health data to inform business decisions and improve performance. Governments are using health data to inform public health policies and optimise health sector spending. Every stakeholder in the health sector is increasingly reliant on health big data, making it an increasingly valuable commodity.

3.1.1 Volume of Health Data

The amount of health data in the world is rapidly increasing. In fact, it is expected to double every 73 days by the year 2020. [1] The volume of health data produced in 2013 was 153 exabytes (1 exabyte is equivalent to 1 billion gigabytes) and an estimated 2,314 exabytes will be produced annually by the year 2020. [2] If these data were stored on a stack of tablet computers, the length of this stack would reach 82,000 miles high—more than a third of the way to the moon. Some estimates put the average amount of health data created by one individual at 750 quadrillion bytes per day, meaning that health data could make up around 30% of the world's daily data production. [3]

Health data are derived from multiple sources with increasing diversification. Major sources of health data include clinical data, genomic data, health internet of things (IOT) data and research data. Most individuals have pieces of health data scattered among various sources, resulting in siloed collections of data. The sum of these health data can be thought of as an individual's 'digital health avatar'. This digital health avatar is an order of magnitude more valuable than the silos of disparate data that comprise it.

3.1.2 Types of Health Data

Health data include any data related to the health, physiology and quality of life for an individual or population. [4] Health data can be broadly divided into clinical, genomic and health IoT categories, each of which are discussed in detail below. Numerous other sources of health data exist, such as data from pharmacies, aged care facilities, clinical research trials, surveys, disease registries and death records. A detailed discussion of these data sources has been excluded for the sake of brevity.

Health data can also be broadly classified as either structured or unstructured. Structured health data are standardised and easily transferable between health information systems. Examples include numerical data, strings, diagnoses, blood-test results and data usually stored in databases. Unstructured health data are neither standardised nor easily transferrable between health information systems; examples include healthcare provider consult notes, medical imaging, audio recordings and physician notes about a patient. In 2013, it was estimated that approximately 60% of health data in the United States were unstructured. [5]

Clinical Data

Clinical data are comprised of content created at healthcare facilities, such as data produced at a point of care at a medical facility, hospital, clinic or practise. The data from these sources are created and held in electronic health records (EHRs) and practise management systems. Other traditional sources include data from personal health records (PHRs) and patient portals. Thousands of software options for these exist around the world. There is also a growing shift towards the use of real-time health systems, which go beyond traditional EHRs to include dynamic IoT and other data integrations and provide a more comprehensive picture of an individual.

Healthcare providers create health data in such forms as notes for a consult, discharge summaries, diagnoses, prescriptions, allergies, family history, social history and physiologic monitoring data. These data commonly include sensitive information such as a patient’s sexual and mental health histories. Clinical data also include investigative results such as lab tests, medical imaging and electrocardiograms, as well as other relevant data like demographic, administrative and medical claims information.

Raw clinical health data usually exist in both structured and unstructured formats. Most subjective data, such as the history of a presenting complaint for the patient’s consult, will take the form of unstructured data. More objective data, such as the findings from a physical examination, prescribing records and diagnoses, are more likely to take the form of structured data. For example, a diagnosis is usually entered into EHRs using a standardised medical classification framework such as the International Statistical Classification of Diseases and Related Health Problems 10 (ICD-10), which is the medical classification list created by the World Health Organisation (WHO) and contains codes for diseases, signs, symptoms, abnormal findings, complaints, social circumstances and external causes of injuries or diseases. The code set in the base classification of the ICD-10 allows for more than 14,400 different codes. [6]

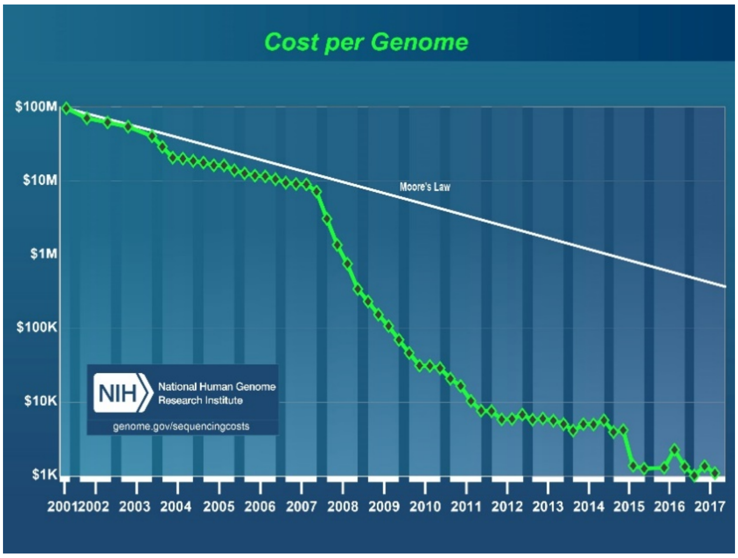
Genomic Data

Genomic data describe an individual's DNA makeup. The DNA sequence of every individual is completely unique and provides a blueprint for how an individual develops from a single cell into a human adult. Genomic data can also reveal an increasingly large amount of information about that individual; they can predict things like what diseases an individual is susceptible to and how he or she will respond to hundreds of different medications and foods. Genomic data are the most important in a greater realm of 'omic data', examples of which include proteomic data (the entire complement of proteins produced by an individual) and microbiome data (the combined genomic data of the microorganisms in an individual's bowel). Genomics is at the helm of the health data revolution and is the foundation of every individual's 'digital health avatar.' As a consequence, it is also the foundation for the future of the health data marketplace, and broader opportunities for big data in healthcare rely on capturing this genomic data opportunity.

There is a consensus among the scientific and business communities that a genomics boom is underway. The emergence of Clustered Regulatory Interspaces Short Palindromic Repeats relying on the protein Cas9 (CRISPR/Cas9) technology has been hailed as one of the biggest discoveries of the 21st century and will lead to the emergence of a plethora of ground-breaking new 'gene therapies'. Indeed, in December 2017, the Food and Drug Administration (FDA) in the United States of America (USA) approved the first-ever gene therapy, Luxturna, for the treatment of an inherited form of vision loss. This milestone is just one example of the potential of genomic data.

Moreover, the shrinking cost of genome sequencing has enormous implications for the widespread adoption of gene sequencing. The first human genome was decoded in 2003 and took 13 years and nearly $2.7

billion in US dollars (USD) to complete. [7] The costs of sequencing a human genome have plummeted, outpacing even the famous Moore's Law. Historically, the high costs of sequencing meant that it was affordable only to the wealthy. For example, in 2011 Steve Jobs spent $100,000 (USD) on whole-genome sequencing as part of his fight against cancer. Today, however, unprecedented scientific advancement has led to the ability to sequence a whole human genome in a matter of days for only a few hundred dollars.



With the costs of genetic sequencing finally reaching levels that enable mass adoption, the amount of genomic data is set to explode. Less than 0.04% of individuals worldwide have actually had their genomes sequenced to some extent. These facts have led to a ‘gold rush’ as genomic companies look to collect genomic data from around the world. Furthermore, over 96% of the genomes sequenced to date having been derived from Europeans, leading to much higher values placed on non-European genomic data [8]. Several companies with valuations over one billion dollars have emerged in this market over the past few years. This is even though they have only procured very small segments of genomic data (around 0.025% of the total genome of each individual) from their customers. These companies generate hundreds of millions of dollars in revenue by selling anonymised genomic databases.

Health IoT Data

Health IoT data (also called the Internet of Healthcare Things and the Internet of Medical Things) are comprised of a vast ecosystem of applications that encompass nearly every industry and institution. Health IoT includes both direct-to-consumer applications and industrial applications comprised of several broad and overlapping categories. Personal wellness devices are perhaps the most obvious; 113 million wearable personal wellness devices were sold in 2017 and the trend for consumer ownership of, and engagement with, their health continues to grow. [9] Fitness wearables (e.g., FitBit, Sensoria) create fitness data such as the number of steps walked during a day and the number of calories burned. Non-wearable sleep monitoring devices (e.g., Beddit, NovaSom) track sleep data such as the number of hours of total sleep and the portion of sleep spent in deep 'rapid eye movement' (REM) sleep. Consumer home-monitoring devices (e.g., GlucoVista, Qardio) track vital physiologic data such as blood glucose levels and blood pressure. Even baby monitoring devices (e.g., Sproutling, Mimo) have emerged to provide infant sleep data and behavioural insights.

Remote health monitoring is another major area of health IoT. Healthcare delivery is increasingly shifting away from hospitals and clinics to private environments such as patients' homes. This shift is being enabled by a variety of health IoT products and is being driven by data supporting improved outcomes for patients and significant healthcare expenditure savings. In November 2017, the FDA approved the world's first 'smart pill': Abilify MyCite is a pill with a sensor that digitally tracks whether patients have ingested their medication. [10] Objective medication adherence and compliance data from such IoT technologies are critical for accurate research. Remote health monitoring IoT companies (e.g., VitalConnect, EarlySense) produce clinical-grade biometric sensors that record real-time clinical data such as vital signs.

Other important health IoT devices include those that connect 'smart' healthcare facilities with remote operations and controls. In fact, it is expected that by 2019, 87% of healthcare organisations will have adopted

IoT technology. [11] Examples include companies like TeleTracking and Awarepoint, which create organisational-level health data. Other companies, like Augmedix and AdhereTech, create clinical efficiencies, while companies like iClinic and CareClix enable telemedicine. These and other similar companies produce large amounts of clinical data that were traditionally limited to EHRs.

3.1.3 Health Data Considerations

Health data are undoubtedly among the most sensitive of all data. As technological innovations in the health sector continue, health data security must remain a top priority for organisations of all sizes. More electronic data are available now than ever before, and efficient and secure ways to manage all this data are needed.

Privacy

One of the main concerns with data privacy in general is the growing system complexity that comes with interoperability. [12] As health systems continue to spend billions of dollars on interoperability, health data move in a myriad of opaque ways, outside any clear consumer understanding. For example, Google DeepMind is an artificial intelligence (AI) company that seeks to collect health data and act as an 'intermediary between the structured health data and any other parties'. [13] Independent reviewers have issued warnings about its potential privacy concerns, and in 2016 DeepMind was found to be in violation of patient privacy laws by gathering 1.6 million patient records at the Royal Hospital of London to test their app.

At the same time, there are ongoing questions about the reliability of 'anonymised' data. Studies have shown that people listed in anonymous genetic data stores could be unmasked by matching their data to a sample of their DNA. One study even found that it is possible to discover the identities of people who participate in genetic research studies by cross-referencing their data with publicly available information on the internet. [14]

Security

In May 2017, a cyber-attack on the British National Health Service brought hospital trusts to their knees, with some disruptions lasting weeks. While individuals' health data were not compromised, the scale and ease of the attack provided a cautionary tale against the centralised storage of sensitive health data. The incident highlighted the importance of securing health data and of protecting the privacy of individuals. Other recent incidents like the Cambridge Analytica scandal and Equifax hacks have also emphasised the likelihood that similar events in the health data realm could potentially occur in the future.

Regulation

Throughout the world, various health data regulations exist to safeguard health information. One of the most comprehensive and well known of these is the US Health Insurance Portability and Accountability Act (HIPAA) of 1996, whose Title II covers protected health information (PHI) across three groups of events: data privacy, data security and breach notification. [15] PHI includes information such as a patient's name, date of birth, medical conditions and any care provided to that individual. Data privacy covers who can access and use the PHI and describes different relevant stakeholder entities. Data security covers how the PHI is safeguarded and includes standards of implementation specifications for health data security. Lastly, breach notification covers the action steps that must be taken after an entity fails to protect PHI. HIPAA also implemented the National Provider Identifier Standard, which requires every healthcare entity to have a unique 'national provider identifier' number, and the Transactions and Code Sets Standard, which requires healthcare entities to follow a standardised framework for electronic data interchange to process claims.

More recently, in May 2018 the European Union implemented the General Data Protection Regulation (GDPR). Although it is not specific to health data, GDPR 'applies to all companies processing the personal data of data subjects residing in the Union, regardless of the company's location'. [16]

Health data are, of course, included as one component of personal data. GDPR enforces new rules and regulations, including requiring sufficient customer consent to process data and stipulating that consent must be given in an intelligible and easily accessible form, with the intended purpose of data processing attached to that consent. Furthermore, users must be able to easily and voluntarily withdraw their consent at any time. The penalties for not complying with GDPR are significant and include a fine of 4% of total worldwide annual turnover. [17] While existing companies have largely failed to comply so far, many start-ups are now building their privacy structures with GDPR in mind. For example, many start-ups are using 'privacy by design' models where data protection is included from the onset of system design.

Ethics

The ethical considerations of data collection and analysis are hotly contested, and nowhere more so than in the health sector, where data are deeply personal. Health data, however, are not limited to the health sector. Interestingly, organisations like Facebook that hold various health data are not subject to health-specific privacy regulations like HIPAA. [18] This issue has recently caused much heated debate.

One of the major ethical considerations around health data is who owns and, by extension, should control health data. This has never been more important as data owners are becoming increasingly aware of the value and importance of controlling their data. The sharing and use of data to create downstream monetary value is no longer a secret and raises questions around equitable remuneration for health data. It would seem fair that individuals should be remunerated proportionately to the value created by those using their data. For example, if a new drug is discovered from research and development involving, to whatever extent, an individual's data, that individual should reap a proportional and fair remuneration from the drug's profits. This contrasts with today's mainstream practise wherein large companies collect, curate and sell health data for large sums of money, often with outcome-based licensing deals that can deliver even more revenue than the original data's purchase price.

An overreliance on health data for decision-making can also lead to general ethical concerns about treating individuals as data points. On the other hand, some researchers have argued that offering health data to improve models is an ethical duty that benefits communities, and thus consent is not an important consideration in light of the public good. [19] This approach views the subject from the perspective of how the potential research could serve people, rather than focusing on the practicality or logistics of gaining consent.

The use of algorithmic learning in the context of health data has been an area of great potential but has not come without potential concerns. For example, AI plays a large role in health data analysis but carries technological and ethical quandaries of selection bias, wherein certain groups are disfavoured. [20] In addition, traditional AI systems are often proprietary, rendering investigation into potential biases impossible. Moreover, the quality of the data has a massive impact on outcome, which is often unaccounted for in academic studies that focus on theory rather than implementation. In practice, imbalanced datasets result in biased results. For example, minority groups, who are typically underrepresented in training data, receive worse treatment as a result of models that only generalise to certain groups.

Researchers at Stanford University have delineated three main types of bias that can affect health data: 'human bias; bias that is introduced by design; and bias in the ways health care systems use the data'. Design bias may occur when an algorithm experiences conflicts between different objectives, such as insurance rates, ability to pay, and saving money—resulting in different outcomes between patient groups. Furthermore, models trained from health data should not be the final answer to any problem. Otherwise, models may lead to self-fulfilling prophecies, an issue across all AI use cases, not just health models.

3.1.4 Health Data Market

The traditional business model of contemporary health data companies is to collect, curate and sell health data to buyers. This model generates billions of dollars in annual revenue for health data companies. Furthermore, licensing deals involving remuneration for revenues generated from the utilisation of data can generate even greater returns. Indeed, the health data brokerage industry alone is estimated to be worth around $47 billion (USD). The pharmaceutical industry alone has a combined market size of nearly $1 trillion (USD) and an annual research and development budget of over $150 billion (USD). A single new drug created from the use of health data can generate billions of dollars in revenue. Similar potential exists for breakthroughs in the multi-trillion-dollar personal-wellness industry. The health IoT market is growing rapidly, and the 'smart healthcare' market is expected to reach $169.3 billion (USD) by 2020, with a significant portion expected to be from remote monitoring. [21] Other multi-billion-dollar industries affected directly by health data include the medtech, biotech and healthcare analytics markets. Insights from health data are crucial to informing policy and maximising profits within the multi-trillion-dollar insurance market. Health data are also crucial for informing government health spending decisions and policies. In fact, it is estimated that around 20% of the healthcare spending in OECD countries is wasted and that the healthcare industry could benefit from an estimated $300 billion (USD) in annual savings by making better use of big data. [22], [23] Therefore, health big data transcends several multi-billion and multi-trillion-dollar industries within the health sector.

3.2 BLOCKCHAIN AND DISTRIBUTED LEDGER TECHNOLOGIES

Distributed ledgers are databases in which transactions are recorded, shared and synchronised dynamically across a network of 'nodes'. These nodes are comprised of computers participating in the distributed ledger network and can be spread across multiple physical sites and institutions. Distributed ledgers enable transactions within a network to be visible to all nodes in the network. This contrasts with traditional ledgers which keep this information centralised. In this way, distributed ledgers provide transparent trust in the network and remove the need for trusted intermediaries.

Blockchain technology builds on decentralised ledger technology. Blockchains record transactions on the distributed ledger in the form of constantly growing 'blocks' of completed transactions. These blocks are recorded and added to the ledger in chronological order to form a 'chain' of blocks. Blockchain technology organises data so that transactions can be verified and recorded through the consensus of all parties involved. For the health sector, this means that any data entered into a computer system can have each transaction or entry validated.

The second generation (and beyond) of blockchain technology, such as Ethereum, allows for self-executing replicated code, and many companies have taken advantage of this to create decentralised services. It is possible to digitise, code and insert practically any document into the blockchain, creating an indelible record of 'smart contracts' that cannot be changed. As the cryptographic properties of larger blockchains (e.g., Ethereum) make blocks virtually un-hackable and immutable, user data privacy can be guaranteed. However, this requires that the initial smart contracts used to facilitate transactions be highly secure as well, without vulnerabilities to known attacks, and audits are necessary to ensure that best practices are in place.

Tokenisation is the process of converting rights to real world assets into a digital token on a blockchain. [24] Blockchain tokens take many different forms but are generally classified under the categories of currency tokens, service utility tokens, security tokens and asset tokens.

3.2.1 Blockchain Health Data Companies

In response to the multitude of security, privacy, regulatory and ethical concerns surrounding health data, a new wave of blockchain companies have emerged. These health data vendors offer solutions to a market where highly sensitive health data are traditionally aggregated and stored on central servers, with no architectural guarantee of privacy or security. This approach gives consumers the responsibility of maintaining control over their own health data. In this way, organisations are forced to come to the level of consumer data, rather than creating complex systems to facilitate data being brought to various organisations. Furthermore, user incentivisation is enhanced because fair compensation for health data is woven into the fabric of the product.

Generally, this new wave of companies aims to create user-centric solutions where users can own and monetise their health data across verticals. Dozens of blockchain health data vendors and brokers have already entered this market at an early stage; thousands more are expected to emerge over the coming years as room for competition in the space continues to be vast. Numerous genomic, clinical and health IoT DApps will emerge, catering to the needs of their local and niche target markets. The barriers to entry for these DApps will be hugely decreased by the existence of blockchain and decentralised ledger platforms catering to their specific needs. Sustainable and fair monetisation of their user base's data will also be greatly simplified by connecting users to a decentralised health data marketplace.

4 PROBLEM ANALYSIS

The current health data market is riddled with problems affecting every stakeholder involved. Distributed ledger and blockchain technologies can provide solutions to many of these problems. HealthDex is focused on tackling two key problem areas that, if solved, would enable a giant leap forward for the health data market. The first problem is the lack of a blockchain platform uniquely tailored to the demands of health data DApps. The second problem is the lack of a decentralised marketplace for trading health data.

4.1 The Platform Problem

As discussed in previous sections, a new generation of health data blockchain companies are rapidly entering the health data market. These companies provide consumer facing products in the form of decentralised apps (DApps). Furthermore, many traditional health data companies are increasingly considering a shift to decentralisation as the benefits begin to stack up. As data protection and decentralisation become more important, health data consumers are beginning to change their consumption behaviours. Currently, there is no end-to-end blockchain platform catering to the unique requirements of enterprise health data vendors (i.e., health data DApps) and consumers. Such requirements include:

- Decentralised storage according to the unique regulatory requirements of health data in different jurisdictions (e.g., HIPAA compliance);
- Data processing according to the unique characteristics and standards of health data (e.g., ICD-10 coding);
- Sustainable data monetisation, especially in relation to static health data (e.g., genomic data);
- Encrypted data analysis, especially in relation to highly sensitive personal health data (e.g., mental health records).

The availability of such a platform would enable a plethora of both new and old health data companies to rapidly build and scale health data DApps. Without the complexities of maintaining a decentralised network, these companies could concentrate on what they do best: growing their user base and providing a great user experience.

4.2 The Marketplace Problem

Thousands of companies and organisations around the world are involved in buying and selling health data. Whether these are blockchain companies or not, the exchanges typically take place through individually-brokered deals made between vendors and consumers. This process is typically slow and inefficient, requiring companies to have large business development overheads. Additionally, these barriers result in the overall suboptimal utilisation of the world's health data. Currently, there is no enterprise health data marketplace where all data providers, consumers and other marketplace participants can easily search, connect, trade and utilise health data. There are numerous reasons for this, but two critical factors stand out. Firstly, the relatively smaller volume of data and the number of stakeholders meant that the need for such a platform was not as urgent. Secondly, the creation of an adequate platform was significantly restricted by technological limitations. Recent changes to both these factors, along with several recent macro trends, have made this an opportune time for the creation of such a marketplace.

5 HEALTHDEX: A COMPLETE SOLUTION

HealthDex solves each of the problems discussed in the previous section and proposes additional opportunities for the health data market. HealthDex is an enterprise health data blockchain platform and decentralised exchange that offers its platform as a service to health data blockchain companies wanting to build efficient health data DApps. The HealthDex decentralised exchange is an enterprise-level (or 'wholesale') health data marketplace. HealthDex is positioning itself to be the world's leading commodities trading platform for health data, enabling enterprises to empower individuals with complete ownership, control and fair remuneration for their health data.

5.1 STAKEHOLDERS

Participants on the HealthDex platform would first need to register and become verified before being provided with access to build and trade on the platform. The verification process will ensure that marketplace participants can legitimately participate on the platform. It will also issue credentials to each user that authorise them to participate in whatever capacity is relevant to them. HealthDex participants fall under four broad categories: 1) data vendor, 2) data consumer, 3) data curator and 4) other. Once authenticated, users will be able to log in to a tailored dashboard to search, connect, trade and utilise health data on the platform. A deeper discussion of the various categories of HealthDex participants follows.

5.1.1 Data Vendors

Data vendors include health data contributors, providers and sellers. They can be categorised as either individual or enterprise data vendors. Individual data vendors are individuals who have a direct HealthDex account, in which case HealthDex is the custodian of their health data. Enterprise data vendors are companies and organisations that have created an account with HealthDex and act as the custodians for the health data of their user base, which in turn consists of individuals who have created accounts with the enterprise in question. Companies that broker and sell clinical, genomic and IoT health data are examples of enterprise data vendors. They can be further classified as either HealthDex DApps (i.e., blockchain enterprises which have built their DApp on the HealthDex platform), Non-HealthDex DApps (i.e., blockchain enterprises that have plugged into the HealthDex marketplace but have built their DApp on another platform), or traditional data vendors (i.e., non-blockchain enterprises that have plugged into the HealthDex marketplace).

5.1.2 Data Curators

Data curators are registered entities and consist of either individuals or enterprises that are incentivised to curate health data on the HealthDex platform. Examples include companies specialising in data curation as well as freelancing data scientists and other individuals. Data curators will convert and combine raw and unorganised data, from across different data sources, individuals and enterprises, into organised data assets which can then be sold in the HealthDex marketplace. These curated data products are much more readily usable than the raw data and are therefore worth significantly more. By creating these data products, data curators are entitled to a share of the revenues generated from the sale of such products, along with the data vendors from whom data is derived.

5.1.3 Data Consumers

Data consumers are individuals or enterprises on the HealthDex platform that purchase data assets to use for analyses and insights. Data consumers include a spectrum of industries and institutions within the health sector. Examples include pharmaceutical, biotech and other life sciences companies, as well as insurance companies, government health departments and academic institutions. Data consumers will consume health data on the HealthDex platform either directly or indirectly. Direct data consumption involves purchasing and analysing data on the HealthDex platform via application containerisation technology, while indirect data consumption involves training encrypted artificial intelligence models on encrypted data using fully homomorphic encryption and federated learning (discussed in more detail below).

5.1.4 Other Marketplace Participants

A number of other actors will participate in the HealthDex platform. For example, data brokers are individuals and enterprises that connect data consumers with relevant raw and curated data. Data validators are individuals and enterprises involved in validating the authenticity and accuracy of data on the HealthDex platform.

5.1.5 Nodes

HealthDex uses a public blockchain platform for transactions to be verified on the marketplace. Therefore, continued function and maintenance relies on rewarding important platform stakeholders, called nodes. These include public blockchain nodes that provide computational power for blockchain transaction verification and, where applicable, AI nodes providing graphics processing unit (GPU) based computational power for federated learning.

5.2 HEALTHDEX TOKENS

HealthDex nurtures a ‘sharing economy’, preventing the concentration of data ownership within profit-driven entities. To this end, HealthDex uses a token system, wherein tokens function as both service utility tokens and a form of currency on the HealthDex platform. These tokens function as incentives, shaping a ‘tokenomics’ system that expands the pool of people able to participate financially and helps keep the HealthDex system focused on human-centric growth. Importantly, this type of platform creates a new business model that emphasises user-buyer inclusion and transparency, thereby encouraging responsible, reliable data usage and transactions. HealthDex Tokens (HDTs) are therefore intended to provide a straightforward, transparent token issuance model that benefits users while supporting the development of the platform. Additional reasons why the HealthDex platform requires a unique token system to function optimally are discussed in more detail below.

5.2.1 Node Remuneration

The various nodes on the HealthDex platform must be financially incentivised to carry out their roles, and remuneration of these nodes is based on predetermined mathematical formulations. Nodes in the blockchain ecosystem are fundamentally incentivised by earning tokens which they believe will increase significantly in value over time. Therefore, any linear remuneration for services (such as USD remuneration) will not be perceived by nodes as providing a significant return on their investment.

5.2.2 Speed and Micropayments

The tokenised service layer ensures frictionless marketplace activity through instantaneous remuneration of all stakeholders. Traditional platforms involve simple linear transactions and/or are not dependent on a rapid exchange of value. The HealthDex platform involves complex multi-party transaction and remuneration. As a dynamic data marketplace, the platform demands the ability to transfer value to multiple relevant stakeholders instantaneously. If a traditional fiat model of exchange, for example using credit card or PayPal payments, were utilised, the platform would function extremely slowly and inefficiently.

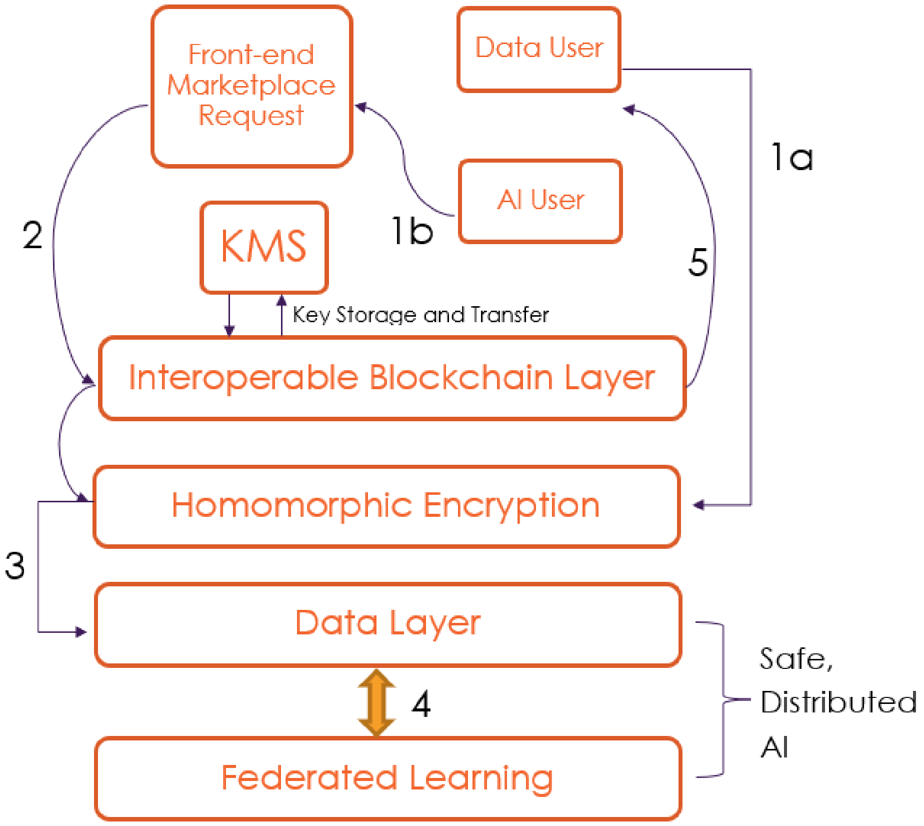
5.2.3 Market Proxy

HealthDex aims to be the world's leading commodities trading platform for health data. As such, the HDT and associated metrics will act as a proxy for measuring and comparing the market values of various health data points and, indeed, the market at large. The health data marketplace is unique and dynamic, and its performance must be analysable without the biases inherent in generic tokens (e.g., Ether). The unique and intricate workings of the platform also render the use of a third-party token impractical. Therefore, not only must the HealthDex platform have a tokenised service layer, but it must also operate using unique HDTs that enable an efficient, scalable and reliable platform.

5.2.4 Other Benefits

Other benefits of the HDT include the ability to provide extra incentives for early adopters, thereby increasing the potential of a network effect, as well as the ability to ensure appropriate functioning within relevant regulatory frameworks (e.g., HIPAA compliance), which other tokens cannot achieve due to their lack of regulation.

5.3 TECHNICAL COMPONENTS



5.3.1 Front End

The front end consists of the interface for the HealthDex marketplace, in which data vendors and data providers securely register and connect. The marketplace links directly to Smart Contracts, enabling blockchain transactions from an off-chain marketplace. Users can easily search the marketplace to find compatible data and models using a search engine based on Lucene to give users the ability to perform full text search and a better user experience.

5.3.2 Entity Resolution

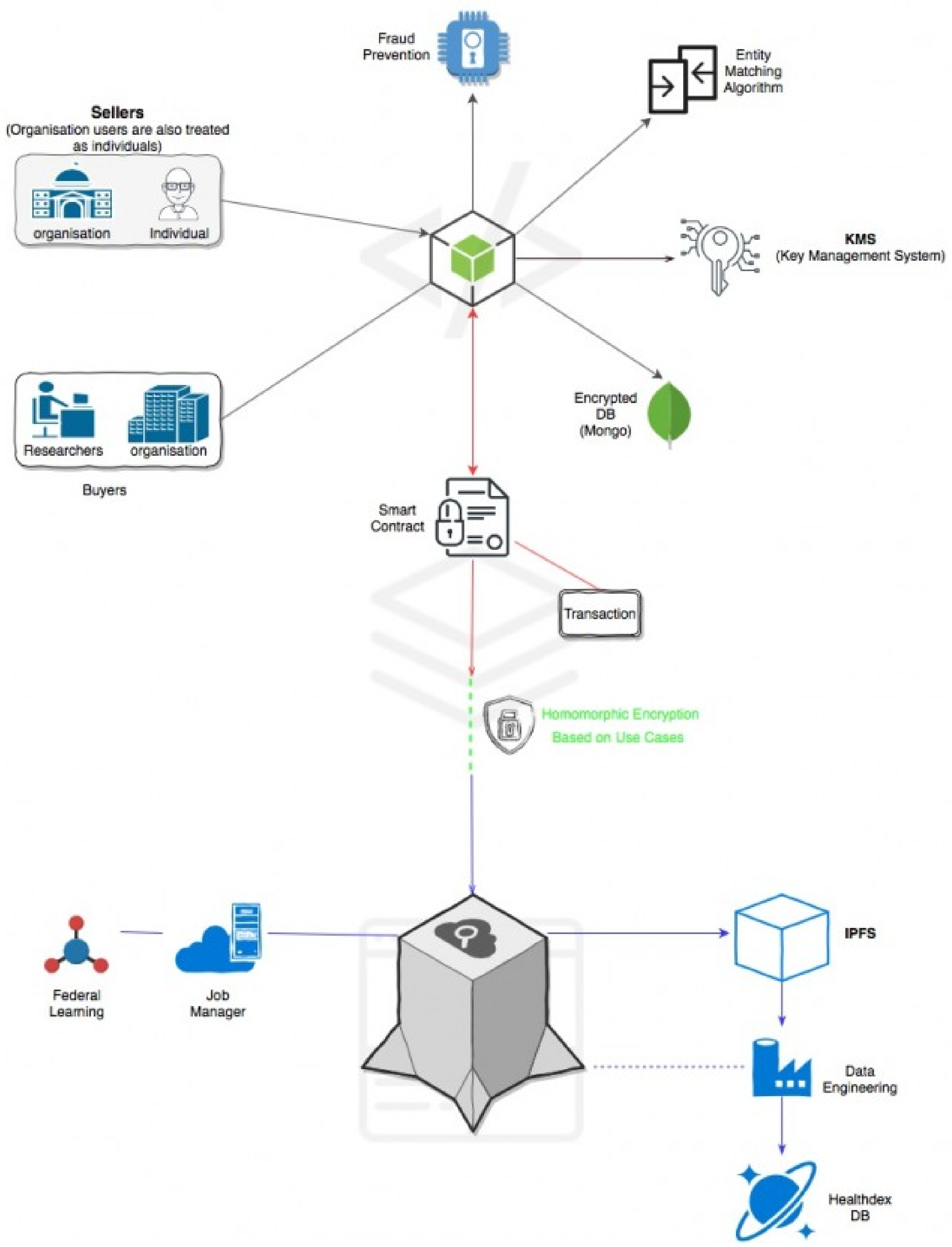
HealthDex implements entity resolution algorithms to settle merge-purge for individuals across disparate sources and create a more functional dataset. By making use of AI based algorithms based on fuzzy and probabilistic matching, HealthDex can unify structured and unstructured general data and create a unified data asset for individuals. This approach enables us to have a singular digital identity of individuals associated with data stored in HealthDex irrespective of the fact whether the data comes directly through the individuals or through a vendor.

5.3.3 Decentralised Key Management System and Encryption

Encryption plays a crucial role in our architecture, and the efficient management of encryption keys is important in order to perform analytics on encrypted data and to share time-based access to data. HealthDex uses fully homomorphic encryption whenever computations are needed on encrypted datasets; however, it is computationally expensive. Therefore, a time-based re-encryption based on symmetric encryption is used for other marketplace use cases, such as when data is provisioned for limited timed applications (e.g., data curation transactions) and for exploratory data analysis.

Key storage is provided by a decentralised key management system (KMS). A central KMS, such as that offered by HashiCorp, Amazon or any other third party, would add an additional external central source to the architecture that we hope to avoid since HealthDex aims to achieve a decentralised architecture. Since the data nodes are private permissioned and distributed, it is vital for HealthDex to have a decentralised KMS, thus ensuring that no excessive amount of trust is placed on any single party. This set-up also ensures that in the unlikely event of a decentralised KMS system being compromised, the data nodes will still be secured from malicious actors.

Healthdex - Off-chain Marketplace



5.3.4 Fully Homomorphic Encryption (FHE)

Encryption is valuable and necessary in the context of a marketplace for health data and AI, as private user data must remain confidential and especially in light of newly-enforced regulations such as the GDPR. Homomorphic encryption allows computations to be carried out on ciphertexts, generating an encrypted result that, when decrypted, is equivalent to the result of operations performed on the plaintexts. In other words, neither HealthDex nor the AI users of our platform have to see user data in order to perform certain computations on the data.

For our use case, we implement encryption on the models and associated gradients (i.e. model parameters) resulting from user training but not directly on the user data. Since one end of our platform enables decentralised AI, nodes need the ability to train AI models without compromising the intellectual property of that model (e.g., by being able to access an unencrypted model). Therefore, models are homomorphically encrypted before users are able to download the model. Users may then train the model to improve accuracy, improving iteratively from one node to the next; this is known as federated learning, and is described in greater detail in the next section.

HealthDex uses a custom algorithm developed by our researchers to implement fully homomorphic encryption (FHE), thus allowing for fully encrypted computations. FHE means that additive and multiplicative operations are both allowed; this contrasts with partial encryption libraries such as Paillier in Python and the Paillier cryptosystem, which only allow for additive operations. The allowance of full computations, as opposed to partial, is necessary for training valuable AI health models. HealthDex FHE uses the Fan-Vercauteren (FV) homomorphic encryption scheme, which is ported from Brakerski's Learning With Errors (LWE) scheme. In essence, LWE is a machine-learning problem that is conjectured to be difficult to solve. Our FHE algorithm is based on cryptographic assumptions used in Fan-Vercauteren scheme and is also complemented by LWE assumptions to achieve security alongside efficient data retrieval protocols.

Since homomorphic encryption can be very computationally expensive, we have provided insights on its performance that clearly show that our implementation is both a feasible and practical method of encryption. The FHE implementation demonstrates high throughput and accuracy; although a single node may have longer training times, the same process repeats across many nodes with no added cost. In the HealthDex context, companies may perform AI computations on user data in the form of ciphertext, allowing users to retain total privacy. Moreover, since FHE can potentially be achieved using only one cloud server, it is more economically feasible than other encrypted computing methods such as Multi-Party Computation (MPC) or hardware-based solutions such as an Intel SGX. FHE is therefore the most appropriate method for encrypting health data from both an economic and a privacy perspective.

5.3.5 Federated Learning

Federated learning is a machine-learning framework that uses datasets distributed across multiple users and entities. Implementing federated learning allows data owners to bring AI models to their level and to train on their own nodes, rather than uploading their data without any assurance of personal privacy. In other words, federated learning allows HealthDex to decouple the machine learning architectures from user data, thereby securing user privacy. Privacy is further secured using homomorphic encryption.

Due to the need to safeguard privacy, training on homomorphically encrypted health data is a strong use case for federated learning: models providing insights are improved while data vendors are empowered with payment for their data. Furthermore, since HealthDex is enterprise-facing, federated learning can be deployed using fewer machines than if individual data owners were targeted, resulting in a clearer technological stack without the need for sophisticated training scheduling. In the domain of big data, federated learning also brings the critical advantages of reducing latency (by minimising the number of round trips the data take) and reducing power consumption (by minimising processing power to the actual requirements of the distributed nodes).

The following is a generalisation of federated learning which we have adapted to our specific use case:

1. Select a subset of clients to download the current model;
2. Each client in the subset (point 1) updates the model based on their own local data;
3. Clients send the model updates to the server;
4. The central server aggregates the models to create a single improved model.

The machine- and federated learning component of HealthDex implements sketched updates which entail a full model update and then compresses it before sending it to the server. We chose this general approach, as opposed to structured updates where one would directly learn an update from a restricted space, for the sake of user simplicity and transparency.

After importing homomorphically encrypted vendor data, the following logic is applied with the objective of making use of the whole training set to improve upon a model that can be trained locally at each data node:

1. The decentralised KMS issues a re-encryption public/private key-pair for FHE data;
2. Organisational clients, such as hospitals, are given the public key, following the protocol:

   > Repeat until convergence: Organisation 1 computes a gradient, encrypts it with FHE and sends it to the next organisation, and so on, until the final organisation passes the overall sum to the server in point 1;

3. The server, which owns the private key, decrypts the gradient of the whole training set to form a final, more accurate model.

The democratic nature of our architecture, which enables anyone with health data to freely contribute to training AI models, helps mitigate selection bias compared to data coming from randomised control trials. Traditional health data sources are known to disfavour certain groups, while our platform attracts groups that might otherwise be excluded. Furthermore, traditional AI systems are often proprietary, so investigation into potential biases becomes impossible. With the HealthDex platform, initial models are stored on public Smart Contracts, enabling transparency.

To reduce overfitting, we will also apply the “early stopping” technique, a method that prevents any given model to be too fine-grained. Meanwhile, we recognise that as the number of available datasets *K* increases over time, the computational demand will not remain constant. Thus, we plan to optimise the distance function using multi-threading that has been previously tested (Trout & Olson 2009). In particular, a parallel loop will be used whenever possible to maximise computational efficiency:

Algorithm: User Matching by Unsupervised Learning-Like Approach

```
Let $m$ denote the total number of users and data sets
Recycle $D$ as the optimisation objective
Initialise set $C = \varnothing$
Parallel for $j = 1$ to $m$,
C := append(C, argmin(D))
Return (argmin©, ©(C))
```

The final object returned, (©mi©), min(C)), can be mathematically viewed as a tuple mapping X -> f(X) .

Upon dataset selection and model training through federated learning, a distributed machine learning framework, the Smart Contract order is executed, and remuneration is released to the data owner. By default, individual units of the learning framework will attempt to optimise the mean squared error (MSE), as most machine-learning algorithms would. However, we also give users the option to optimise alternative error metrics, including mean absolute error (MAD) and binary/multiclass cross entropy (CE):

$$MSE \ = \frac{1}{n} \sum_{j=1}^{m} \left(\hat{y}_j - y_j \right)^2$$

$$MAD = \frac{1}{n}\sum_{j=1}^{m} |\ \hat{y}_j - y_j\ |$$

$$CE = -\frac{1}{n}\sum_{j=1}^{m} y_j\ Log\ \frac{y_j}{y_j + \hat{y}_j}$$

This enables our federated learning framework to generalise well to a variety of models, including more complex algorithms that optimise for alternative metrics.

5.3.6 Application Containerisation

HealthDex understands that not every data analytics and analysis can be performed using federated learning on fully homomorphic encrypted data, which is why for select groups of HealthDex participants and use cases, secure and encrypted containers will be made available for the users to perform their data analysis.

5.3.7 Blockchain Layer

The front end connects to the blockchain layer, as users execute Smart Contracts from the front end to trade data and models. Users are rewarded with HDTs for sharing data with AI users. The Smart Contracts that regulate transactions on the HealthDex platform allow users to buy and sell health data through two main contracts. The contracts send and receive payment in the form of HDT, enabling the circulation of utility and creating an internal economy. As the Smart Contracts reside in a public blockchain, transactions within the HealthDex marketplace provide transparency and auditability—useful features in any transparent health data ecosystem.

In order to understand the Proof of Stake (PoS) Blockchain security of HealthDex, let us first explore Proof of Work (PoW). PoW secures bitcoin, and is a method of determining if a new block being created is valid or invalid. This, in turn, is determined by the amount of "work" done to solve a problem and generate a solution. The network agrees upon a task to solve with adjustable difficulty, and whoever (of the "miners") solves this task first gets to find a new block and then begin the search for the next block. Ultimately, the longest chain has had the most “work” done, so it can be trusted. This system incentives miners to get more hardware, using copious amounts of electricity, and potentially centralizing the system to few miners.

The most popular solution to the aforementioned problems is Proof of Stake (PoS), which secures the network by high stakes. In the HealthDex PoS Blockchain ecosystem, every user has a chance to find the newest block, proportional to the amount of stake in internal currency they commit (HealthDex Token). Fake transactions would effectively make the currency you stake in become worthless, so users are disincentivized to do so, and the network is secured without the computational inefficiencies of PoW. As new problems arise from PoS, HealthDex will likely switch to a Cosmos model in the future.

Our implementation of the HealthDex platform uses a two-tiered hybrid blockchain system, architected as a public (staked) and unpermissioned system for transactions and contract management and a private permissioned system for health data regulation–compliant storage. This means that permission is required in order to read the encrypted data. This is especially useful considering the privacy concerns over health data, which are described in the next section.

5.3.8 Data Layer

Data on the HealthDex platform are stored in one of two primary sets of nodes: encrypted databases and a permissioned network of enterprise nodes. Using the hybrid blockchain, a consortium of blockchain health companies constitute the network of private storage nodes.

Public Data

Metadata from users are stored in an encrypted database within the front-end application. These data include users’ wallet ID, address, ethnicity, age, gender, and ‘flag’ (a Boolean value signifying whether the user is sharing their data). These pieces of data, though not exhaustive, enable user privacy alongside sufficient information for data buyers to search on our platform.

PERMISSIONED ENTERPRISE NODES

The private data are the actual health data that individuals sell on the platform. These data are secured in private and permissioned enterprise nodes, as they must undergo a verification process to contribute data, including meeting HIPAA compliance standards, and are only open to partner firms within the health alliance. In the absence of such companies, the data would be stored in HealthDex nodes. Furthermore, structured or prepared data is stored in the HealthDex database (HealthDexDB), which is a uniquely tailored database providing efficient data processing for health data vendors and which has blockchain characteristics such as decentralisation and immutability, based on BigchainDB (which itself rests on MongoDB).

Naturally, the data cannot be stored on the blockchain, as this would quickly lead to overwhelming inefficiency as well as prohibitive cost. Our hybrid system of permissioned data nodes and a public blockchain for contracts is chosen for two main reasons. First, avoiding public nodes allows HealthDex to remove the threat of future unpredictable quantum attacks that could decrypt data help on a public IPFS. Second, a permissioned data system is necessary for meeting current regulatory compliance.

IPFS is a hypermedia distribution protocol that creates a content-addressable way to store and share data in a distributed file system. This enables HealthDex users to search and share distributed health data and is also useful for maintaining authenticated health data, as it serialises the data to create a 'Merkle DAG', which is a cryptographically-authenticated data structure. The off-chain marketplace connects to metadata in IPFS nodes to select data that fit the profile of a user search. These nodes are not stored on public IPFS, but rather on private permissioned HealthDex servers. An IPFS object specifically contains two fields: data and links. The data are unstructured binary data, while the links are an array of link structures that connect to other IPFS objects. The link structure has three data fields: name, hash, and size. Name and hash are of critical importance to our architecture.

DATA ENGINEERING

The data engineering layer lies beneath the data layer in our architecture and is where the data is formatted and prepared to ensure that the data adheres to industry standard data models. The algorithms for data preparation are deployed on HealthDexDB nodes. Data engineering can be characterised as using software engineering principles to manipulate data prior to higher-level data science paradigms, such as deep learning. Conceptually, data engineering practices will follow the common ETL pattern: extract, transform and load. Extract simply means bringing the data from our data layer to the algorithm-containing nodes. The core of our data engineering, i.e. transforming the data, then takes place. This involves business intelligence and analytics applications, such as aggregating, merging and filtering data to create actionable data. Finally, the data are loaded and transported to the federated learning layer, where model training occurs.

Our data engineering processes will vary according to data type in order to maximise the breadth of users in our platform. For example, electronic health records (EHR) and personal health records (PHR) are governed by the Health Level-7 (HL7) standards in the USA to provide data communicability between organisations. We implement different sets of data engineering algorithms for each health data type and its corresponding data standard. The functional specification that HL7 offers for EHR and PHR includes a highly specific and detailed framework known as 'Version 2 messaging', more commonly referred to as 'Pipehat'. These standards detail the encoding syntax for the data that our algorithms will implement. Many different data formatting standards are defined under the Clinical Data Interchange Standards Consortium (CDISC), which supports medical research of any type. These standards cover XML, data tabulation, laboratory data, operational data, and many more. Implementing a comprehensive data engineering layer with the standards discussed above will enable HealthDex AI users to tackle more ambitious projects and use complex models without spending too much time on unifying disjoint datasets.

HEALTHDEXDB AND DATA CONSENSUS

HealthDex uses Tendermint for the consensus layer in our architecture for HealthDexDB. Every HealthDexDB node contains an individual MongoDB database, and communication between these databases across nodes is completed using Tendermint consensus. Tendermint is an alternative consensus mechanism and is the first Practical Byzantine Fault Tolerance (PBFT) algorithm in a Proof of Stake (PoS) blockchain. Using BFT means that a single trusted entity is not needed for consensus and the HealthDexDB nodes communicate using Tendermint wiring protocol. The data asset registry concept (discussed later in the whitepaper) forms the basis of why a separate consensus is needed on the data layer.

5.4 OTHER FEATURES

5.4.1 Entity Matching Algorithms and Blended Data

Disparate, disconnected and isolated data are almost always less useful than a functional dataset where multiple types of data about an entity are available in one place to help with the analysis. The HealthDex platform provides blended data by matching the entities where data resides in different data sources with numerous blockchain and traditional data companies. This is accomplished using probabilistic and fuzzy-matching proprietary algorithms for the public data while ensuring that the privacy of the private data is not compromised. This enables organisations that specialise in one or more types of health data, for example genomics data, to benefit from any other data that the individuals are plugging into the marketplace.

5.4.2 Data Management and Ownership

Upon joining HealthDex, individual data contributors and users from organisational data-contributing companies are registered in the Entity Analytics application, where algorithms are used to assign unique IDs to each individual based on their public data. This process ensures that the next time the same entity's data gets plugged into the ecosystem through another organisation, the entity is not duplicated. The public data of entities, along with their profiles, are stored in an encrypted database in the

Marketplace Application, complying with the GDPR and HIPAA compliance requirements.

5.4.3 Data Security and Encryption

HealthDex values user privacy and takes related rules and regulations very seriously. To this end, we have built our platform to accommodate GDPR from the start. We take GDPR into account as it ‘applies to all companies processing the personal data of data subjects residing in the Union, regardless of the company’s location’, and we process data internationally.

In response to new rules and regulations, HealthDex requires sufficient customer consent to process data, and the request for consent is given in an intelligible and easily accessible form with the purpose for data processing attached to that consent. Furthermore, we provide users with the ability to easily and voluntarily withdraw their consent at any time. HealthDex also implements Privacy by Design, and we have included data protection from the onset of the design of our systems. Most importantly, HealthDex users provide sensitive data as ‘opt-in by design’, in that the purpose of our platform for users is to provide data in a data marketplace. Finally, our system is built upon verifying compliance and trust for regulatory bodies through Smart Contract technology, enabling existing institutions to adopt the HealthDex platform and help shape the future.

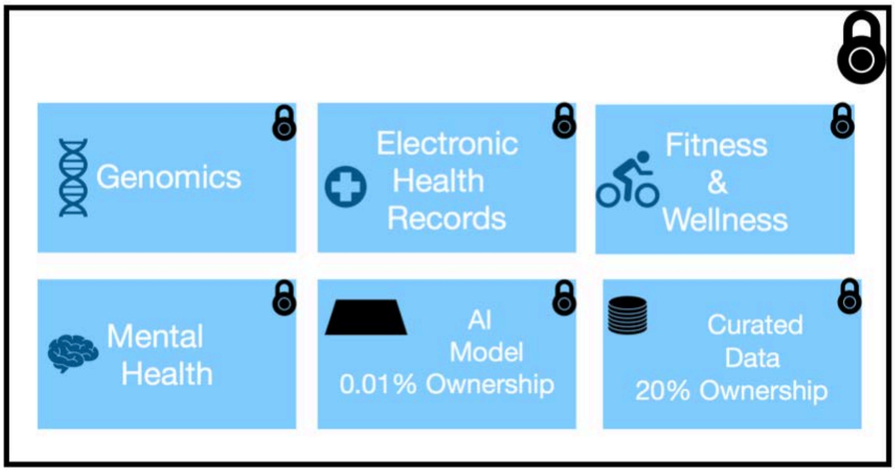
HealthDex divides each individual's data into two parts. The first is their public data, which consist of their publicly identifiable information such as name, gender, date of birth, addresses and ethnicity. The second is their private data, which are their actual health data; time-based access to this is sold over the marketplace. The public data are used by researchers and data buyers to identify the groups of users whose data they would like to buy. The private data's metadata (binary flags) are used to identify whether the private data are present or not. The actual private data are encrypted and stored in private permissioned distributed nodes and are accessed through re-encryption time-based private keys, which only the individuals control.

5.4.4 Fair Distribution

The HealthDex public blockchain for transactions is a distributed ledger which contains all marketplace transactions. When blended data is traded, there are multiple parties involved who have stakes in the commodity (data assets) that is transacted. The public ledger enables anyone to explore the transactions and the distribution of funds from each transaction. The marketplace will have multiple entities involved with every transaction, which is why a public ledger is extremely important to show a fair and public distribution of the transactions. This includes remuneration but also others involved as intermediaries, such as data curators.

5.4.5 Data as a Transferable Asset

All data on the HealthDex platform are treated as assets. Every time an individual's data is crowd sourced by a curator or researcher to create a new data artefact, the individual's data asset registry records a stake in the curated data. Every time the curated data is monetised; the individual gets a stake in that trade. A similar ownership structure is used for ownership in AI models created and rendered on HealthDex platform by data scientists who crowd source data. The data assets in this registry are also considered to be transferable, whereby an entity can transfer the ownership and rights of their registry to another entity (e.g., family trust or charity). This will ensure indefinite and ongoing utilisation and monetisation of data assets if desired.



5.4.6 Fraud Prevention

Fraud prevention will be necessary to ensure the validity and accuracy of the shared data. HealthDex incentivises a specific category of marketplace participants, termed data validators, to carry out this role. In this way, data will be graded according to its authenticity and accuracy. Users found to be falsifying data will be banned from the HealthDex platform.

6 REVENUE MODELS

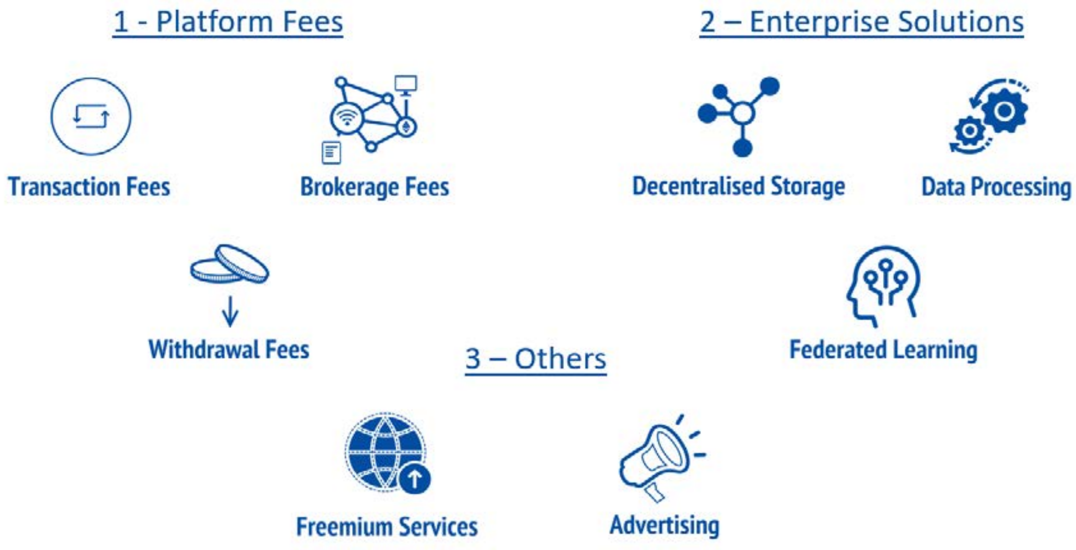
HealthDex's revenues are primarily generated from commission fees. Because of this, HealthDex is incentivised to maximise the number of market participants and trading volume on the HealthDex platform. These fees are charged on all transactions, withdrawals and brokerage services that occur over the HealthDex decentralised exchange. All transactions over the platform will incur a transaction fee payable by transacting parties at a fixed percentage rate, charged in HDT. Similarly, all withdrawals of HDT from HealthDex wallets to external wallets will incur a withdrawal fee at a fixed percentage rate charged in HDT.

In addition, HealthDex will broker outcome-based licensing deals using Smart Contracts to enable data contributors to profit long-term from the use of their data. For example, a pharmaceutical company will share profits generated from a novel drug discovery to data contributors who helped make it possible. All transactions related to historically-agreed Smart Contracts will incur brokerage fees at a fixed percentage rate charged in HDT. Transaction and withdrawal fees on the HealthDex platform are analogous to online market exchanges. For example, the 24-hour trade volume in the cryptocurrency market surpassed $50 billion USD in December 2017 but this is still far behind the foreign-exchange market, which has a 24-hour trade volume of over $5 trillion USD. [25] Nonetheless, the largest five cryptocurrency exchanges each have over $1 billion USD per 24-hour trading volume (Data: CoinMarketCap.com). Based on a 0.1% commission, the daily revenue of such exchanges exceeds $10 million USD. Given this, one of the two most important metrics for the HealthDex marketplace will be the daily trade volume: the higher this is, the higher

HealthDex's commission revenue will be. This metric will be closely tied to the number of active participants on the HealthDex platform.

The second major revenue model for HealthDex is through the provision of enterprise solutions and HealthDex's platform-as-a-service model. Health data DApps wishing to use HealthDex's services will be charged according to the exact services they wish to purchase. This will take the form of a 'freemium' subscription model whereby basic services are provided without charge and additional features must be purchased if and when required. For example, for health data vendors and brokers, HealthDex's distributed storage and uniquely tailored health data processing features would be most likely to appeal, whereas for health data and AI consumers, HealthDex's federated learning and encrypted AI-as-a-service offering would likely be of more interest.

Other forms of revenue that HealthDex will generate include those from consulting services, with a focus on helping new blockchain health data companies. The HealthDex Venture Fund will actively invest in and assist promising health data blockchain start-ups looking to build on the HealthDex platform. Finally, and following the creation of significant network effect, HealthDex will tap into the $4 billion USD medical advertising and clinical trial recruitment markets.



7 REFERENCES

[1] IBM, “Healthcare’s Data Dilemma: a Blessing or a Curse?,” 2018 (accessed).

[2] IDC, “The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things,” 2014.

[3] E. Fry and S. Mukherjee, “Tech's Next Big Wave: Big Data Meets Biology,” 2018. [Online]. Available: http://www.fortune.com/2018/03/19/big-data-digital-health-tech/.

[4] Farlex, “Health Data,” 2018 (accessed). [Online]. Available: https://medical-dictionary.thefreedictionary.com/health+data.

[5] Datamark, “Unstructured Data in Electronic Health Record (EHR) Systems: Challenges and Solutions,” 2013.

[6] World Health Organization (WHO), “International Statistical Classification of Diseases and Related Health Problems (ICD-10),” 2010. [Online]. Available: http://apps.who.int/classifications/icd10/browse/Content/statichtml/ICD10Volume2_en_2010.pdf.

[7] C. Gyles, “The DNA revolution,” *The Canadian Veterinary Journal,* vol. 49, no. 8, 2007.

[8] A. B. Popejoy and S. M. Fullerton, “Genomics is failing on diversity,” *Nature,* 2016.

[9] CBInsights, “Google in Healthcare,” 2018 (accessed). [Online]. Available: https://www.cbinsights.com/reports/CB-Insights_Google-Healthcare-Briefing.pdf?utm_campaign=google-health_2018-02&utm_medium=email&_hsenc=p2ANqtz--ZHdawcHtrDrZJbnSkKZOy7xd5GUzTzDKC8HXrTrsVlJ61AMZuaa0jdyzlN776IRCrDEeAeURIpFdrGnbUGtcl1boqWw&_hsmi=61339017&ut.

[10] FDA, “FDA approves pill with sensor that digitally tracks if patients have ingested their medication,” 2017. [Online]. Available: https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm584933.htm.

[11] Aruba: a Hewlett Packard Enterprise Company, “IoT Heading for Mass Adoption by 2019 Driven by Better-Than-Expected Business Results,” 2017. [Online]. Available: http://news.arubanetworks.com/press-release/arubanetworks/iot-heading-mass-adoption-2019-driven-better-expected-business-results.

[12] J. Thune , L. Alexander , P. Roberts, R. Burr and M. Enzi, “Where Is HITECH’s $35 Billion Dollar Investment Going?,” 2015. [Online]. Available: https://www.healthaffairs.org/do/10.1377/hblog20150304.045199/full/.

[13] N. Lomas, “UK report warns DeepMind Health could gain ‘excessive monopoly power’,” 2018. [Online]. Available: https://techcrunch.com/2018/06/15/uk-report-warns-deepmind-health-could-gain-excessive-monopoly-power/.

[14] G. M, A. L. McGuire , D. Golan , E. Halperin and Y. Erlich, “Identifying personal genomes by surname inference,” *Science 339,* vol. 339, 2013.

[15] US Dept. of HHS, “Summary of the HIPAA Privacy Rule,” [Online]. Available: https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html.

[16] GDPR, “General Data Protection Regulation (GDPR),” 2018 (accessed). [Online]. Available: www.eugdpr.org.

[17] M. Drolet, “GDPR fines: How much will non-compliance cost you?,” [Online]. Available: https://www.csoonline.com/article/3234685/data-protection/gdpr-fines-how-much-will-non-compliance-cost-you.html.

[18] F. Donovan, “Health Data Privacy Rears Its Head at Facebook Hearing,” 2018. [Online]. Available: https://healthitsecurity.com/news/health-data-privacy-rears-its-head-at-facebook-hearing.

[19] M. Wood, “Consent and the Ethical Duty to Participate in Health Data Research,” 2018. [Online]. Available: https://blogs.bmj.com/medical-ethics/2018/01/25/consent-and-the-ethical-duty-to-participate-in-health-data-research/.

[20] P. Hannon, “Researchers say use of artificial intelligence in medicine raises ethical questions,” 2018. [Online]. Available: https://med.stanford.edu/news/all-news/2018/03/researchers-say-use-of-ai-in-medicine-raises-ethical-questions.html.

[21] Technavio, Global Smart Healthcare Market 2016-2020, 2016.

[22] OECD, “Tackling Wasteful Spending on Health,” OECD Publishing, Paris, 2017.

[23] IBM, “How Big Data Equals Untapped Opportunities and Savings,” 2018 (accessed). [Online]. Available: https://www.ibm.com/watson/infographic/discovery/big-data-savings/.

[24] A. Cameron-Huff, “How Tokenization Is Putting Real-World Assets on Blockchains,” 2017. [Online]. Available: https://www.nasdaq.com/article/how-tokenization-is-putting-real-world-assets-on-blockchains-cm767952.

[25] O. Williams-Grut, “The cryptocurrency market is now doing the same daily volume as the New York Stock Exchange,” 2017. [Online]. Available: http://markets.businessinsider.com/currencies/news/daily-cryptocurrency-volumes-vs-stock-market-volumes-2017-12-1011680451.

[26] “IBM Healthcare: IBM's Data Dilemma; A Blessing Or A Curse | IBM,” [Online]. Available: https://www.ibm.com/industries/healthcare/datadilemma.

[27] M. Felleisen, “Accumulators,” in *How to Design Programs*, Second ed., Boston, Northeastern University, 2018 (accessed), p. Epub.

[28] O. Trott and A. J. Olson, “AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithrea,” *Journal of Computational Chemistry,* 2009.

[29] J. Fan and F. Vercauteren, “Somewhat Practical Fully Homomorphic Encryption,” *Cryptology ePrint Archive,* no. 144, 2012.

[30] Z. Brakerski, A. Langlois, C. Peikert, O. Regev and D. Stehlé, “Classical Hardness of Learning with Errors,” *arxiv.*