Fysical: A Decentralized Location Data Market

February 1, 2018

MOTIVATION

Over the last five years, the location data market has grown substantially, resulting in an influx of buyers and suppliers entering the market. There are now thousands of data suppliers that sell some form of location data. The demand for location data has seen the most growth in the marketing industry, but there is burgeoning demand in financial services, real-estate, and artificial intelligence.

The influx of location data buyers and suppliers has created benefits, notably in greater awareness of location data use cases, but the rush of companies has also led to mass inefficiencies in the market.

Today, centralized data exchanges facilitate much of the location data trade from suppliers to buyers in this market. These exchanges charge exorbitant fees that range from 30% - 50% on average. Moreover, their position as an intermediary has resulted in a net value loss for both data buyers and data suppliers. For data buyers, centralized exchanges provide no transparency on how data is generated and who generated it: understanding both is required to evaluate whether data is high-quality, accurate, and privacy compliant. For data suppliers, centralized exchanges don't let suppliers control their data, the exchanges blur the lines on data usage and ownership, and their ambiguity on quality causes the data to be constantly mispriced.

The inefficiency in centralized exchanges and lack of transparency on data provenance (origin), data usage, and privacy compliance, has stifled adoption for many industries that could benefit from buying or selling location data. This has become a problem, especially for companies working with artificial intelligence and machine learning which depend on buying high-quality data.

Fysical has fixed these problems inherent in this booming location data market, and we're excited to share this solution with the world:

Fysical is a decentralized location data market protocol that removes middle-men and provides for the transparent and efficient exchange of location data.

Fysical is the leading data market for urban planners, retailers, academic researchers, investors, and many more to better understand how humans move through the physical world, enabling them to make better decisions.

CONTENTS

1.0 LOCATION DATA

1.1 What is location data?

Location data, which may also be referred to as "human movement data," is any piece of data that relates to where a person is currently, has been historically, or is predicted to be in the future. Some of this data is explicitly generated by an individual, such as by checking into a location on an app.[1] Other data is generated implicitly, such as by tapping your bus pass at a bus stop.[2]

Below is a sample of some data sources that produce location data:

- Social media check-ins
- Daily lifelogs
- Web surveys
- Wi-Fi connectivity (IP address) via laptop, tablet, or phone
- Receipt data
- Credit card data
- First-party transactions to stores/events
- Location-Stamped Images
- Satellite Imagery
- Mobile app location services
- Visible Light Communication
- Ultrasonic audio communication
- Visual Foot Traffic Counters
- Connected car signals (Bluetooth, GPS)
- Tickets via train, plane, or bus
- Gym check-ins
- Restaurant reservations
- Calendar meeting location entries

Furthermore, new datasets can also come into existence by joining some of the above primary sources of location data together.

1.2 How is location data used today?

Today, location data is utilized by a wide range of companies, organizations, governments, and machines to make important decisions. These decisions can range from simple, where a retailer adjusts its staff based on the number of visitors to its stores, to complex, where emergency first responders are quickly dispatched after a natural disaster according to the distribution of people in affected areas. Below is an example of some industries that rely on location data:

---

[1] "What Is Implicit Data? - Definition from WhatIs.com." WhatIs.com, Dec. 2012, whatis.techtarget.com/definition/implicit-data.

[2] "What Is Explicit Data? - Definition from WhatIs.com." WhatIs.com, Dec. 2012, whatis.techtarget.com/definition/explicit-data.

- Advertising Technology: Targeting based on location and geo-behavioral shopping habits
- Artificial Intelligence (AI): Research and discover patterns in location data for automation and greater operational efficiency
- Augmented Reality (AR): Validate the locations of players in real-world mobile games
- Autonomous Vehicles: Guidance for these vehicles based on traffic and congestion; Goldman Sachs predicts that the market for guidance for autonomous cars alone will grow from around $2.2 billion in 2020 to $24.5 billion by 2050[3]
- Financial Services: Real-time foot traffic analysis of publicly traded retailers to infer sales volumes, home appraisal, and fraud prevention by banks
- Government & Smart City: Public transit service route optimization, improved disaster response, defense, and resource management
- Healthcare: Fitness tracker data for health analysis and geo-alerts for wandering dementia patients
- Insurance: Helps insurers to automatically adjust rates or verify claims based on consumer physical activity or motor vehicle usage
- Logistics: Improved routing and advanced detection of bottlenecks
- Real Estate: Site selection based foot traffic evaluation of surrounding and similar areas
- Retail: In-store mapping and guidance for product discovery
- Telecommunications: Smarter decisions about cable, towers, and service coverage by examining population density trends

1.3 What is the future of location data?

Location data will eventually be self-generated by sensors and smart devices that are deployed throughout the world.[4] These sensors will be a part of the Internet of Things and will document all information about the locations and people that they detect. This is happening already, largely because of the growth in mobile devices that sit in your pocket. Mobile devices are extremely advanced and have a plethora of high-accuracy sensors that can report back movement data[5], such as your precise location, indoor floorplans, your altitude, speed, bearing, and your position relative to Bluetooth-enabled devices or Smart TVs.

1.4 How do companies buy location data?

Today, data buyers typically buy data from centralized data exchanges. These exchanges have done the work of aggregating datasets from data suppliers, who are willing to sell their data, which the exchanges then sell on their behalf. The exchanges pay data suppliers with a monthly revenue share in the range of 50% - 70% of the total revenue

---

[3] "The Battle for Territory in Digital Cartography." The Economist, The Economist Newspaper, 8 June 2017, www.economist.com/news/business/21723173-not-all-roads-lead-google-maps-battle-territory-digital-cartography.

[4] "Geo IoT Technologies, Services, and Applications Market Outlook: Positioning, Proximity, Location Data and Analytics 2016 - 2021." Research and Markets, Mind Commerce, Aug. 2016, www.researchandmarkets.com/research/f84z3b/geo_iot.

[5] "Sensors Overview." Android Developers, Google LLC, 25 Oct. 2017, developer.android.com/guide/topics/sensors/sensors_overview.html.

received from data buyers. The entire process of supplying data to receiving payment for data normally takes 90 to 120 days. Considering the novelty and importance of this market, it remains a fairly conventional buyer-seller process—slow, inefficient, and with high upcharges for ineffective middlemen.

Much of this inefficiency stems from the lack of transparency in the quality and source of the data in the market. Purchasers of data do not pay until they know the full value of what they have received, which can take two to three months. This is because it is only after they receive the data that they have the time to employ engineering resources to comb over the results and filter out duplicate, fraudulent, or faulty data. If data buyers could verify the reputation of data suppliers and compare dataset overlap before making a purchase, then they would know a dataset's exact worth and feel comfortable in paying for it in real time or even in advance.

Below is an example of how location data is exchanged today:

1. Day 1 – Day 30:
    a. Data Supplier transfers the daily dataset to Centralized Exchange.
    b. Centralized Exchange passes on daily dataset to end Data Buyer.
    c. Data Buyer uses data.
    d. The process repeats daily until Day 30 when it restarts for the next month.
2. Day 45 – Day 60:
    a. Centralized Exchange invoices Data Buyer for data already used.
3. Day 60 – Day 90:
    a. Data Buyer pays Centralized Exchange.
4. Day 90 – Day 120:
    a. Centralized Exchange takes a revenue share (40% of total revenue)
    b. Centralized Exchange pays Data Supplier (60% of total revenue)

2.0 PROBLEMS AND SOLUTIONS

There are a few core problems with the current model for buying and selling location data. We have outlined these problems and how Fysical solves them below.

2.1 Data Buyer Problems and Solutions

1. **Problem:** There is zero transparency on data provenance (origin). Centralized exchanges provide no accessible record of facts concerning the data's source or owner. This creates a cobweb of duplicate data purchases, and labor and time spent analyzing data overlap by data buyer engineering teams.

   Solution: Fysical keeps a permanent, immutable audit trail of atomic[6] exchanges of FYS tokens for datasets. These transactions are factual records about a dataset's provenance, and it is freely accessible by any data buyer who wishes to validate a dataset's source.

2. **Problem:** Data buyers have no permanent record that a dataset was created in a privacy-compliant manner with the authorization of end consumers.

   Solution: Fysical enables suppliers to append privacy compliance documentation, such as privacy policies and user opt-in messages, that are dated with timestamps for every dataset that is published to Fysical. Data buyers now confidently buy datasets that are authorized for use by the end consumers, and they can now audit the privacy compliance of datasets at any time, even those datasets purchased in the past.

3. **Problem:** There is no easy way to determine that a dataset is accurate or high-quality. Data buyers have to run a manual analysis on each dataset to understand its quality before putting the data to use. This can take months and often requires looking at multiple datasets at once.

   Solution: A pool of FYS tokens are to be released over time with more tokens rewarded to data suppliers who begin publishing data to Fysical sooner. The rewards token pool acts as an incentive for data suppliers to provide high-quality, in-demand datasets to data buyers. Furthermore, because Fysical utilizes the Ethereum blockchain when a data supplier submits data to Fysical, it executes a transaction on the Ethereum network, which results in a permanent record on the Ethereum blockchain. These Ethereum transactions incur an execution fee to be paid in Ether. This fee acts as a natural filter, discouraging data suppliers from uploading "spam," fraudulent, and low-quality data.

4. **Problem:** High-value location events are often not available to buyers until they are published in a dataset hours or days later. This delay prohibits the possibility of many high-value use cases that would involve these events.

[6] "Atomicity (Database Systems)." Wikipedia, Wikimedia Foundation, 22 Jan. 2018, en.wikipedia.org/wiki/Atomicity_(database_systems).

Solution: Fysical allows market-driven data-delivery schedules to emerge. Sellers can identify datasets for sale immediately or prior to when they will be available. Buyers can purchase access to these datasets as soon as they are published or even before they are published. This futures-style trading enables subscription purchases and real-time data delivery.

2.2 Data Supplier Problems and Solutions

1. **Problem:** Once data suppliers transfer data to a centralized data exchange, they cannot determine if its use falls within the terms of their contract or their expectations.

   Solution: Fysical provides data suppliers with a factual record of data usage. After data suppliers submit data to Fysical, they can track who bought it and who re-sold it on Fysical.

2. **Problem:** The pricing of data fluctuates, and payments are unreliable. If a data supplier sells to a centralized data exchange, most agreements are for revenue shares, with payment delays that range from 90 to 120 days after the data has been used by the end buyer; in some cases, payment is never made.

   Solution: Fysical makes pricing transparent and payment upfront and reliable. All datasets published to Fysical have decryption keys that allow access to the datasets. The only way these keys can be obtained and utilized to access a dataset is with an FYS token payment that is agreed upon upfront by the data buyer and data supplier. This means that for data buyers to purchase a dataset, they must pay the agreed upon token amount. Data suppliers will no longer be forced to accept payment 90 to 120 days late or when it's opportune for data buyers.

3. **Problem:** Businesses that may have valuable data cannot easily test market demand for their data because the barriers to entry for new entrants is extremely high and there is a lack of trust in data buyers and centralized exchanges. This means that valuable and potentially innovative datasets remain locked away.

   Solution: Fysical allows anyone to easily publish a dataset, which can then be searched for and purchased by potential data buyers. This lowers the barrier to entry for businesses that are new to the data trade to test the market demand for a dataset, resulting in new, unique, and useful datasets entering the market. Furthermore, Fysical promotes trust from data suppliers because it is built on the blockchain which means all transactions are permanently recorded, and no one entity controls the trade of data.

2.3 Data Ecosystem Problems and Solutions

1. **Problem:** There are no trusted sources of reputations for data suppliers or data buyers.

Solution: Fysical makes datasets and historical transactions reliable and also freely accessible and verifiable by anyone at any time. This means data buyers can better determine whether a data supplier has a good reputation and quality data based on their historical transactions and datasets, quickly and easily.

2. **Problem:** There is no standard data format for location data. There is no standard schema for the exchange of location data. For both data buyers and suppliers, this means customization is required for every product and business transaction to be successful.

   Solution: Fysical allows de facto standards to emerge for data schema, meta-data schema, and delivery mechanisms. Buyers and sellers can acquire and put data to use faster by using the same data formatters, parsers, and transmitters with several partners instead of developing custom mechanisms for each relationship. Furthermore, data can be more easily searched when it's in a standardized format, enabling data buyers to find new and useful datasets quickly.

3. **Problem:** Due to a lack of reliable reputations of market participants and standards for data trade, agreements to exchange data are too slow. The time for data buyers to evaluate potential data suppliers can be well over three months.

   Solution: Fysical is a reliable and trustless source for reputations of both data buyers and data suppliers and allows for a de facto standard to develop in the location data trade. Both can increase the speed at which data trade agreements are executed, making the ecosystem more efficient and enabling faster growth and innovation.

3.0 FYSICAL

3.1 Model

Fysical is a location data market. There are two core parties within Fysical: those who supply location data and those who buy location data. Both the suppliers and the buyers transact with FYS tokens.

Currently, most datasets on Fysical consist of anonymous and aggregated GPS and sensor information from mobile devices across the globe.

New types of data are constantly being added to Fysical, including receipt data, first-party transaction data, credit card data, Wi-Fi access-point data, or satellite imagery. These datasets enhance Fysical's data offering and open up new use cases. New datasets are also crucial to providing accurate location data. That is, combining many signals or relative positions can help to reach a stronger consensus than just one signal that indicates where people are in space and time.

The data made available for sale on Fysical is owned by the supplier, whether they publish it using tools such as a software development kit (SDK), an application programming interface (API) supplied by a third-party, or whether they publish it independently and directly to the blockchain. The suppliers always receive payment in FYS tokens when their data is purchased by a data buyer, and the suppliers can also receive additional tokens in the form of rewards when they contribute data.

Many current data suppliers are mobile app publishers that naturally create anonymous 'exhaust' data, such as GPS location signals, when people use their apps. This exhaust data is generated by individuals simply walking the earth. It's idle and available to be used, similar to how extra storage space on computers is utilized by companies like Storj[7] and Filecoin[8]. The exhaust data is contributed to Fysical so it can be purchased by data buyers.

3.2 Fysical is widely used

Fysical has been in development for three years and has 1,000 data suppliers who contribute more than 15 billion data points from over 10 million mobile devices around the world every month.

Fysical holds a wealth of knowledge about the physical world, human movement patterns, and visited places. For perspective, every 30 days, Fysical takes in more location data records than the complete 9-year history of Foursquare users' location check-ins.[9]

---

[7] "Storj: A Peer-to-Peer Cloud Storage Network." Storj, Storj Labs Inc., 15 Dec. 2016, storj.io/storj.pdf.
[8] "Filecoin: A Decentralized Storage Network." Filecoin, Protocol Labs, 14 Aug. 2017, filecoin.io/filecoin.pdf.
[9] Yeung, Ken. "Foursquare Users Have Checked in over 10 Billion Times." VentureBeat, VentureBeat, 13 Sept. 2016, venturebeat.com/2016/09/13/foursquare-users-have-checked-in-over-10-billion-times/.

4.0 TOKEN MECHANICS

FYS tokens enable transactions between data buyers and data suppliers within the Fysical ecosystem, and the tokens are treated as a utility. FYS tokens and blockchain technology enable atomic transactions[10], which create a reliable way to confirm that a transaction between a data buyer and a data supplier occurred, without the need for a middle-man or escrow service.

The use of FYS tokens also provide numerous advantages over the traditional means of exchanging data; some examples include: verifiable audit trails in the form of factual records of transactions, reliable market participant reputations, the real-time trade of data, a global market rate for location data to enable efficient international data trade, and incentives for data suppliers who are often also data buyers on Fysical.

4.1 The Past: Payment without Tokens and Blockchain

Prior to Fysical, data suppliers were not paid for the datasets that they supplied to data buyers until well after the data had been transferred. The entire process from supplying data to receiving payment for data was normally 90 to 120 days and commonly involved centralized exchanges that acted as middlemen. It is a fairly conventional buyer-seller process—inefficient, and riddle with delays and high fees for ineffective middlemen.

4.2 The Future: Payment with Tokens and Blockchain

FYS tokens and the Fysical blockchain enable atomic exchanges of datasets and FYS tokens between parties. When a buyer finds an encrypted dataset and wishes to acquire the key to decrypt it, the buyer formally proposes an exchange of tokens for the key on the blockchain. The dataset seller accepts that offer by encrypting the key for the buyer and recording that encrypted key on the blockchain. The smart contract ensures that the tokens are transferred to the seller, and the encrypted key is recorded at exactly the same time. Either both operations occur, or neither occurs. At any time in the future, the buyer can read the encrypted key from the blockchain, decrypt it, and use that key to decrypt the dataset. In contrast to how data was formerly traded without Fysical, both of these actions happen simultaneously, and one cannot happen without the other. This largely benefits data suppliers as they can now be paid immediately for the datasets they supply, rather than 90 to 120 days later.

The use of Ethereum smart contracts for atomic transactions allows these two parties to execute this operation simply by trusting the publicly available logic of the smart contract. In contrast, to perform such an exchange with a private centralized provider, both sides must, at the very least, trust the centralized provider.

4.3 Tokens Enable Audit Trails

Because all data published to the Fysical blockchain requires a decryption key and FYS tokens to obtain that decryption key, for the first time in the location data trade, there now exists a technical mechanism that creates an atomic transaction that any entity can

---

[10] "Atomicity (Database Systems)." Wikipedia, Wikimedia Foundation, 22 Jan. 2018, en.wikipedia.org/wiki/Atomicity_(database_systems).

rely on in a publicly accessible, trustless, and decentralized ledger. This enables two powerful use cases that have never existed in the location data trade: 1. Verifiable and publicly accessible audit trails of transactions that have occurred within the ecosystem. 2. Reliable reputations of market participants within the ecosystem. The verifiable and publicly accessible audit trails of transactions enable a reliable reputation system because anyone can now reference the factual history of transactions between data buyers and data suppliers.

Both of these use cases create rapid efficiency within this ecosystem, especially for data buyers who, prior to making a buying decision, can now quickly evaluate potential data suppliers based on their reputations.

Before Fysical, the only way to establish a trustworthy and reliable data trade was through an escrow service, which remains highly expensive, cumbersome, and not publicly available or verifiable.

4.4 Tokens Enable Real-time Trade

The need for real-time location data for automation and artificial intelligence is growing rapidly. Prior to Fysical, these needs were largely unmet in the location data market as most of the data trade is done manually and through large daily transactions. The ability to trade futures on Fysical allows market participants to agree on subscription-style, real-time delivery of datasets, and this enables machines and APIs with access to FYS tokens to make dataset purchases instantly, without the need for a middle-man and with guaranteed dataset delivery. For example, the New York Police Department (NYPD) can build an automated system to increase police presence when foot-traffic levels reach a certain threshold at a protest or assembly. With Fysical, the NYPD's automated system can use FYS tokens to instantly subscribe to and acquire a reliable real-time foot traffic data feed for specific blocks in New York City. Without tokens and blockchain, the NYPD system would not be able to reliably obtain this data in real-time.

4.5 Tokens Enable Efficient International Trade

The international location data trade is growing rapidly. Currently, businesses in the data trade execute via their native currencies, and this opens these businesses to exchange rate risk between currencies.[11] Fysical and FYS tokens change that. For the first time, the value of location data will be expressed as a utility with worth that is determined solely by free market conditions of the location data trade. By utilizing the FYS token to purchase data, international businesses express trade conditions in identical terms and enable transactions that otherwise might not have occurred. For example, a U.S. business that has to buy a dataset in USD on day 0 and resell it for Yuan on day 30 incurs the risk of the exchange rate removing the profit from that deal, but if that business can express both trades in the same currency, then that risk disappears. Accounting for this risk becomes more important as machines and automated systems begin buying location datasets. That is, by expressing trades in the same currency, exchange rate risk doesn't exist, and machines and autonomous agents are free to purchase data at the global rate

[11] "Foreign Exchange Risk." Wikipedia, Wikimedia Foundation, 7 Feb. 2018, en.wikipedia.org/wiki/Foreign_exchange_risk.

when they need it. If machines and autonomous agents had to account for the exchange rate risk at the same time, this could eliminate profits and stifle innovation because they may be stalled from acting or making a purchase that they would have otherwise made with a global rate.

In either case, the use of a common global FYS token for the location data trade eliminates the exchange rate risk and enables more global business relationships and potentials for innovation.

4.6 Token Rewards

There are many use cases that can help incentivize data suppliers to supply data and earn tokens in return. These incentives are designed to supply Fysical with more datasets and new datasets that are in high demand from data buyers across the globe. For data buyers, the rewards help to curate a marketplace, supplied with more robust data for them to purchase.

Below are a few examples of the ways that data suppliers can contribute data for rewards:

Example use case: Polygon Mapping – Incentive for data suppliers to provide geographic location coordinates of a retail store's dimensions and boundaries.
Example use case: Geo Missions – Incentive for data suppliers to provide data in particular countries, cities or stores.
Example use case: Time-sensitive Missions – Incentive to provide X amount of data within the next week.

4.7 Data Suppliers Are Also Data Buyers

Data suppliers are often also data buyers in the location data market. This is because most data suppliers utilize location data for their own products, and buying more data can be an added benefit. For example, many app publishers who supply datasets to Fysical also run applications that require foot traffic data on the popularity of nearby stores. This data is difficult for the app to collect itself, so instead, the app publisher purchases this data from Fysical. In a second example, many data suppliers want to learn more about their users' real-world habits by analyzing additional location data on where their users spend time. They can then improve their app by personalizing the experience to users based on their habits and hobbies. Again, app publishers don't collect enough information on their own users to get a great, holistic understanding of their users' habits; instead, they purchase additional data from Fysical using FYS tokens.

In this light, data suppliers exchange their data for FYS tokens and in turn use those FYS tokens to buy more data, creating a balanced and efficient ecosystem.

5.0 BLOCKCHAIN

5.1 What is blockchain?

Blockchain is technology that offers a digitized, decentralized, immutable public ledger of the exchange of data.[12] Its decentralized nature means that no single person or company controls data entry or its integrity, and its method "chains" together data entries so that they cannot be modified later. Blockchain provides a way for large groups of unrelated companies to jointly keep a secure and reliable record of the exchange of data.

5.2 Why is Fysical built on the blockchain?

For location data to provide value, it needs to be highly accurate, privacy compliant, and it must contain information about how the data was generated and who generated it. Blockchain provides this in a trustworthy and scalable way without the presence of a centralized authority.

For data buyers, blockchain enables audit trails and reveals facts about the origin of data, therefore generating greater confidence in the integrity of the data purchased.

For data suppliers, blockchain provides full transparency on data usage and clear ownership rights on data, thereby creating greater trust in sharing data.

5.3 Why is Fysical built on the Ethereum blockchain?

Ethereum is a globally distributed computer that is run by a network of users around the world who are incentivized in Ether, Ethereum's cryptocurrency, to run resources for protocols like Fysical.[13]

Fysical is built on Ethereum so that it can leverage its blockchain for the decentralization of Fysical and also for utilizing the tokenized aspect of Ethereum to reward suppliers of location data and to facilitate the efficient exchange of location data between data buyers and suppliers.

---

[12] Prableen Bajpai CFA. "Blockchain." Investopedia, 29 Jan. 2018, www.investopedia.com/terms/b/blockchain.asp.

[13] "Ethereum." Wikipedia, Wikimedia Foundation, 7 Feb. 2018, en.wikipedia.org/wiki/Ethereum.

6.0 TECHNICAL DESIGN

6.1 Marketplace Technical Design Philosophy

Since Fysical is designed to improve the efficiency of trade in the markets we encounter, we expect that the characteristics and patterns of future data trade will evolve beyond those of today. To enable this growth, the Fysical smart contract design focuses on the fundamentals of data trade. A diverse set of market participants use these fundamental operations to develop systems to meet varying market conditions.

Among the critical elements of a data market are:

- Atomic exchange of assets
- Long-term, short-term, and future availability of verifiable, immutable datasets of arbitrary size and format
- Formal standards and de facto standards for data schemas, encryption algorithms, compression algorithms, data transport mechanisms, and data storage
- Quality and reputation ratings

Data suppliers, data buyers, trade brokers, rating agencies, and hybrids of all four:

- publish tradable datasets with meta-data describing their properties, including (but not limited to) format and provenance
- find and acquire datasets based on meta-data, availability, and supplier reputations
- aggregate, augment, enrich, normalize, filter, and compress data from Fysical and publish the result to Fysical with a trail of the provenance
- develop user interfaces, programmable interfaces, and automated tools to interact with Fysical

6.2 Ethereum Smart Contract

The market should be built on a public, decentralized, immutable database, supporting atomic transactions. Ethereum, the dominant blockchain-based smart contract platform, supports these effectively.

The ERC20 Token Standard[14] is widely supported public token standard. This means active market participants can easily acquire the Fysical tokens needed to acquire the data critical to their business. In addition, the public support and ease of accessibility for ERC20 tokens lowers the barrier to entry for potential market participants.

The Solidity language for smart contract development[15] has the most exercise in the community of Ethereum developers. The OpenZeppelin project[16] offers a well-exercised implementation of token management compatible with ERC20. Since the logic defined in

---

[14] "ERC20 Token Standard." ERC20 Token Standard - The Ethereum Wiki, 30 Dec. 2017, theethereum.wiki/w/index.php/ERC20_Token_Standard.

[15] "Solidity." Solidity - Solidity 0.4.20 Documentation, 2017, solidity.readthedocs.io/en/develop/.

[16] OpenZeppelin. "OpenZeppelin/Zeppelin-Solidity." GitHub, 26 Jan. 2018, github.com/OpenZeppelin/zeppelin-solidity.

a smart contract is immutable after publication on the Ethereum network, a review by independent auditors ensures that the contract can securely maintain a consistent set of data and allow market participants to trust the fairness of the marketplace.

The Fysical Solidity code is compiled to executable bytecode for publication on the Ethereum network. This Solidity code is published in a public repository along with a reference to the version of the Solidity compiler used to produce this bytecode. Public services like Etherscan offer verification alignment between bytecode found on the Ethereum network and Solidity code in public repositories.[17]

Ethereum, Solidity, ERC20, and OpenZeppelin are the basis for Fysical. Auditors' familiarity and the existing volume of activity using this stack of technologies gives us confidence in the long-term viability of Fysical.

6.3 Dataset Publication and Access

Datasets routinely traded on Fysical are larger than the entire Ethereum blockchain. Storing the content of these datasets directly on the blockchain has never been a practical consideration for the scale expected by our trade partners and competitors. Fysical can hold an immutable, verifiable reference to any dataset referable with a URI[18] outside of Fysical. Practically speaking, datasets are typically served from HTTP servers, FTP servers, cloud storage providers (e.g., Amazon S3, Google Cloud Storage), and BitTorrent. In the future, Fysical will naturally support new protocols and transport mechanisms that can uniquely identify resources using URIs.

Publishing a dataset on Fysical also involves recording the size and a checksum of the data referenced by the URI. The combination of the URI, the size, the checksum algorithm, and the checksum allows data consumers to verify that they have obtained exactly the data referenced on Fysical. The modification of even a single bit in a dataset results in a different checksum using the same checksum algorithm.[19] This same mechanism underpins the security of most blockchain technology.

The cost of an Ethereum transaction depends largely on the volume of data stored on the blockchain during the transaction. Only a small amount of data must be stored on the blockchain to create a verifiable reference to a dataset stored outside the blockchain. The size of the dataset does not affect the amount of data required to store this reference and therefore does not affect the cost of publishing this reference on the blockchain.

Data suppliers may encrypt a dataset before publishing its reference to Fysical. Buyers can download and verify that this encrypted dataset matches the on-chain reference prior to obtaining a key to decrypt the dataset.

---

[17] etherscan.io. “Verify Contract Code.” Ethereum Solidity Contract Source Code Verification | Bytecode | ABI, Etherscan, etherscan.io/verifyContract.

[18] “Uniform Resource Identifiers (URI): Generic Syntax.” Ietf.org, The Internet Society, Aug. 1998, www.ietf.org/rfc/rfc2396.txt.

[19] “Checksum.” Wikipedia, Wikimedia Foundation, 29 Jan. 2018, en.wikipedia.org/wiki/Checksum.

For a variety of reasons, a supplier may stop serving a dataset referenced on Fysical. In some cases, a supplier will find that the cost of hosting a referenced dataset exceeds its value in the market. In other cases, a supplier may simply cease doing business. Fysical cannot monitor or maintain the availability of referenced datasets. Fysical allows suppliers to record their intention to maintain a dataset for a period of time. Market participants, such as brokers and rating agencies, may find value in comparing the actual availability of datasets with the stated intention of the suppliers. Other market participants may find value in mirroring referenced datasets to increase their availability to the market.

In the same way that Fysical supports any data transport mechanism, Fysical allows suppliers to attach any schema for interpreting a dataset at the time of its publication. For example, some data buyers prefer to receive a human-readable, line-delimited set of records with fields separated by commas[20], while others find programmatic advantages in consuming data with a compact binary format defined by Protocol Buffers[21].

A supplier may also choose to attach a description of the data contained in the set. For example, if publishing a set of events all occurring in the United States within the same hour, the supplier may define the hour and the country in the set's description. Other market participants, such as brokers, may develop performant mechanisms to search Fysical for datasets based on these descriptions.

6.4 Dataset Purchasing

The limited performance and cost of transactions on the blockchain do not always suit the nuances of negotiation between the buyer of a dataset and its supplier. However, the permanence of Fysical blockchain records is perfect for recording the clear, formal agreements arising from negotiations. Market participants, including brokers, can develop and re-develop negotiation mechanisms that result in permanent, public agreements on Fysical.

An informal negotiation outside of the blockchain might involve a buyer and a supplier exchanging emails, making phone calls, or an in-person meeting. It might involve a supplier instructing a broker to advertise a fixed Fysical token price for the data and accept offers from any buyer. It might even involve two algorithms negotiating over a new protocol without human intervention. All of these informal methods can result in an offer to purchase that the supplier will accept.

The purchase of a dataset published on Fysical involves two steps. A buyer with a balance of Fysical tokens records a formal offer to purchase a dataset that includes a number of Fysical tokens to transfer to the dataset's supplier. This formal offer can also include arbitrary data relevant to the two parties such as license agreements to be enforced by authorities outside Fysical.

---

[20] "Comma-Separated Values." Wikipedia, Wikimedia Foundation, 5 Feb. 2018, en.wikipedia.org/wiki/Comma-separated_values.
[21] "Protocol Buffers." Wikipedia, Wikimedia Foundation, 6 Feb. 2018, en.wikipedia.org/wiki/Protocol_Buffers.

The dataset supplier executes the second step of formally accepting this offer. If the dataset is encrypted, this acceptance must include a key to decrypt the dataset. To ensure that only the buyer can decrypt this dataset, the decryption key is encrypted, using a public key provided by the buyer in the purchase offer.[22] The Ethereum transaction, recording this acceptance, executes the exchange of Fysical tokens and the decryption key atomically[23].

6.5 Audit Trail

Currently, third-party participants can help to aggregate datasets from a variety of sources. By combining the data from multiple sources, they can provide a cleaner, enriched, and concise dataset, meeting the expectations of data buyers. Though data buyers prefer to work with the data in this new form, they are commonly concerned with the true origin of the data. Because Fysical records a formal history of dataset transactions, by referencing previous Fysical dataset purchases, a data supplier can demonstrate the provenance of a dataset at the time of its publication.

6.6 Bad Data Filter

The Ethereum blockchain requires a fee for executing a network transaction. When a data supplier submits data to Fysical, this is a network transaction on the Ethereum blockchain, and this requires a fee. This means that by using Ethereum, there is a natural bad data filter in the form of the Ethereum network transaction fee, which discourages data suppliers from uploading "spam" or fraudulent data.

6.7 Dataset Availability, Futures, and Subscriptions

Trading partners may arrange the delivery of data in the future in exchange for Fysical tokens in the present. A dataset supplier can create a URI for a dataset to be generated in the future. This URI can be published to Fysical, along with an availability window beginning in the future. At the time of this publication, the supplier would not know the length or checksum of the data, so the buyer would not be able to verify access to the data at the time of purchase. The lack of data at the time of purchase does not prevent the exchange of the key for the data's decryption.

A brokerage could trivially develop the notion of a subscription using Fysical's basic mechanisms for future exchange. For example, a subscriber could instruct a brokerage to purchase any future dataset from a specific supplier, appearing at regular intervals. Naturally, acquiring a future dataset may involve more risk for the buyer than a purchase of a downloaded, verified dataset.

---

22 "Public-Key Cryptography." Wikipedia, Wikimedia Foundation, 5 Feb. 2018, en.wikipedia.org/wiki/Public-key_cryptography.

23 "Atomicity (Database Systems)." Wikipedia, Wikimedia Foundation, 22 Jan. 2018, en.wikipedia.org/wiki/Atomicity_(database_systems).