Datapace

Decentralized data marketplace based on blockchain

Drasko Draskovic
Datapace
drasko@datapace.io

George Saleh
Datapace
george@datapace.io

December 28, 2017

Abstract

Data is the new oil. Every day we create 2.5 quintillion bytes of data[1]. 90 percent of the data in the world today has been created in the last two years alone – and with new devices, sensors and technologies emerging, the data growth rate will likely accelerate even more.

Not only centralized SW giants, but also mobile and network operators and various enterprises that install huge number of devices or any electronic infrastructure are in position to put the sensors in their equipment and collect huge amount of data. Some of this data is perishable - i.e. it must be consumed instantly or it looses value. Some of this data is long-lasting. No mater what kind of data stakeholders collect, they usually have the same problem: how to draw additional profit from this data, beyond it's immediate and obvious purpose[1].

All this data would have value for many parties and can be further monetized. Data collectors could become data sellers, and offer collected data on the specialized data marketplace. On the other hand, data buyers would be interested to browse offered data streams and buy them, then use this data to further process it and/or build new services for their customers.

A global-scale marketplace for IoT sensor data is need. This marketplace is called Datapace.

[1] Take for example a mobile telephony operator. Company like this already owns huge number of network base-stations, gateways and antennas which make the deployed network infrastructure. These network devices are already equipped with big number of telemetry sensors that provide the operating state insights and are used for management and maintenance. Data coming in for this sensors is useful for the operator to keep the network healthy and functional. But beyond this primary purpose, collected data can be extremely useful for other parties - like Smart City municipalities, health institutions or various other businesses. Moreover, because of the density of mobile base-stations and antennas, operators are in unique position to offer for example extremely precise environmental data, which is hard to achieve to even specialized services as it demands significantly expensive HW sensor installation. Similarly, a company that does smart signage could turn public signs into sensing stations with marginal additional costs. With further cost drop of the sensors and appearance of smart dust, even individuals or small enterprises will be capable to collect significant amount of IoT data.

Contents

List of Tables

List of Figures

1 Introduction

Datapace is marketplace for IoT sensor data. But beyond IoT, Datapace marketplace can be used to sell or buy any type of data, independently of it's type or provenance.

Datapace is distributed and decentralized system based on blockchain. Blockchain technology is used for several important purposes in Datapace system:

- To enable tokenization of value (i.e. provide TAS token) and token economy
- To insure data integrity (i.e. to store data hashes and guarantee that data is not tampered with)
- To enable Smart Contract capabilities
- To provide network security via PBFT consensus and immutability and make the system hack-proof

Each of these characteristics of blockchain and how they are leveraged upon in the Datapace system will be explained in more details in the following chapters.

Datapace market place is built with intention to be simple, easy to use and intuitive. Anyone familiar with classic e-store-like web portal should immediately understand how to sell digital assets - in this case the data stream, or how to browse offered data streams and purchase selected data. Simplicity of use opens possibility for mass-market adoption while simplicity of the system provides high quality implementation and better secured and more performant application.

2 Stakeholders

Datapace system is built on private, permissioned blockchain. It uses PBFT algorithm for concensus and state replication, which guarantees high transaction throughput and fast transaction finality (which as a consequence prevents blockchain forking). Because of the nature of PBFT algorithm, the whole system is run by a closed consortium with a known set of validators. Never the less, any entity can potentially request access to the consortium and run a validating node under contractual agreements.

Based on this we can identify following stakeholders of Datapace system:

- Data buyers

- Data sellers
- Validators

2.1 Data Buyers

Data buyers are organizations and individuals that are interested in buying the data. They log into the system and browse the data streams offered for sell, as one would browse items on e-store web site.

Data streams are offered under certain price and can are purchased for TAS tokens.

Data buyer must have a sufficient amount of TAS tokens in his wallet in order to purchase the data. Once data is purchased, data buyer obtains a proxied HTTP URL from which he can consume the data. This URL is unique and temporary - it expires after the lease period for which data was payed for.

2.2 Data Sellers

Data sellers are organizations or individuals that offer the data for sell.

It is responsibility of data seller to provide a valid data source URL and give detailed description of the data stream an it's format (it's JSON schema) - so it can be easily consumed by data buyer. This URL is secret, and it is never reveled to data buyer. It is only temporary proxy URL that is given to data buyer, and it expires after time data was payed for.

Additionally, data seller can provide geolocation data of the stream source, so that it can be queried on the maps.

Data sellers should provide valid data sources. In order to guarantee the validity of the data, Datapace employs several mechanisms - like seller reputation rating and verified IoT gateway HW provisioning, which will be explained in dedicated chapter.

Data sellers obtain TAS tokens in their wallet when the stream that they offered is purchased.

2.3 Validators

Validators are the entities that participate in network infrastructure, i.e. in block validation. Validators are rewarded for their work in TAS tokens.

Because in the phase 1 Datapace is based on private PBFT blockchain, set of validators must be known up-front. Datapace consortium will allow adherence of new members under strict contractual agreements.

In the second phase of development, Datapace validation will be opened to public via novel *Proof-of-Verified-Source* and *Proof-of-Stake* on the Cosmos[2] network.

3 System Architecture

3.1 Description

Datapace is a decentralized application based on the blockchain network with native token of value.

Datapace blockchain is based on Hyperledger Fabric technology, an industrial blockchain implementation with quality guarantees by Linux Foundation and consortium of over 200 companies gathered around the open-source project.

In addition to Hyperledger Fabric, Datapace comes with specialized Smart Contract (chaincode) which implements ERC-20 token (crypto-currency) native to the platform.

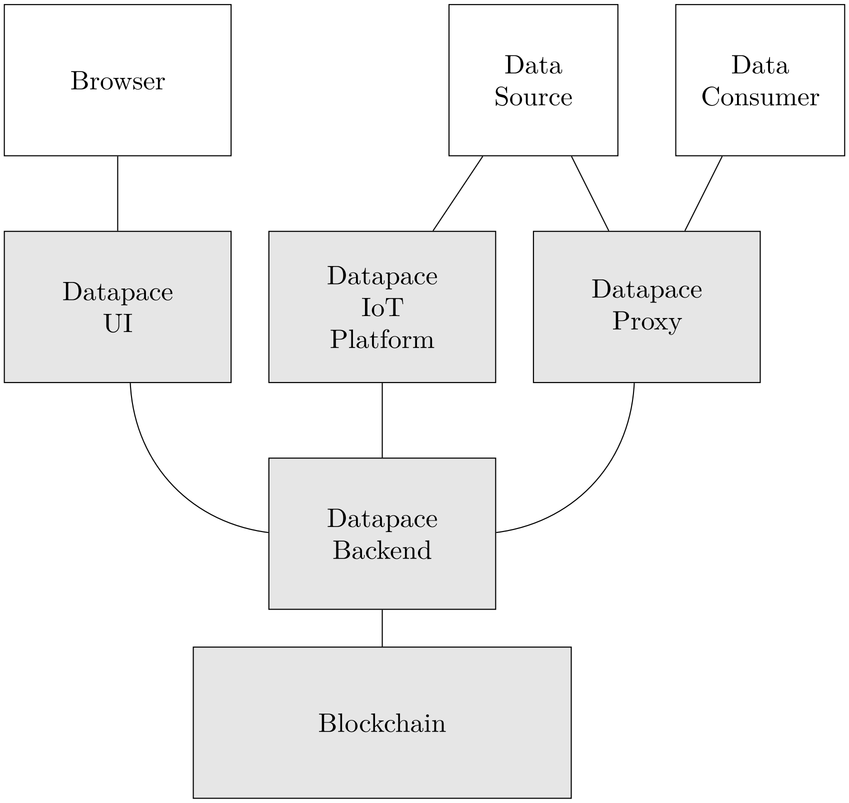

Figure 1: Datapace System Architecture

Hyperledger Fabric uses high-performance PBFT consensus algorithm - it supports thousands of transaction per second at 1000ms latencies. Additionally, an ABC-compliant connector/adapter for Datapace system to incoming Cosmos network will be created in the future. Announced as “Internet of Blockchains”, Cosmos hub will give to Datapace system two very important features: interoperability and additional scalability.

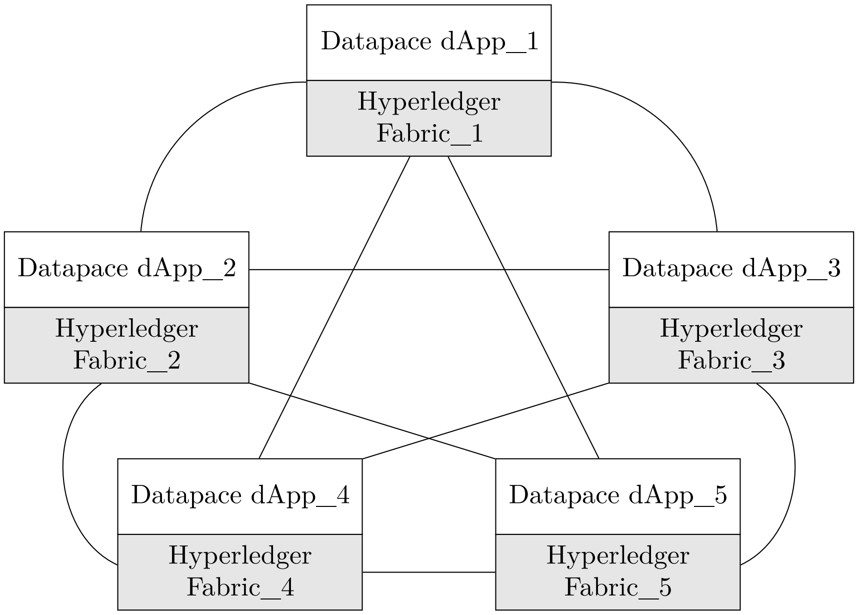

Figure 2: Datapace Blockchain Network

Interoperability is extremely important, as it will enable TAS token to natively flow from Datapace private blockchain into other blockchains connected to the Cosmos hub, thus opening potential for TAS exchange to other crypto-currencies, and vice versa. This will influence token economy and raise the value of the TAS token. Additionally, developed token economy would allow *Proof-of-Stake* consensus to be applied on the top of the Datapace-Hyperledger system and allow opening Datapace validator set participation to the wide public.

Scalability is also important, although, as a consequence of the wise technology choices, Datapace system is already very performant. But “Interent of Blockchains” will enable additional scaling od Datapace chains through sharding[3] using Cosmos zones.

As mentioned before, blockchain technology is used for several important purposes in Datapace system:

- **TAS token**: TAS token is native token of value in Datapace system and is necessary for system operation and functioning. It will be explained in

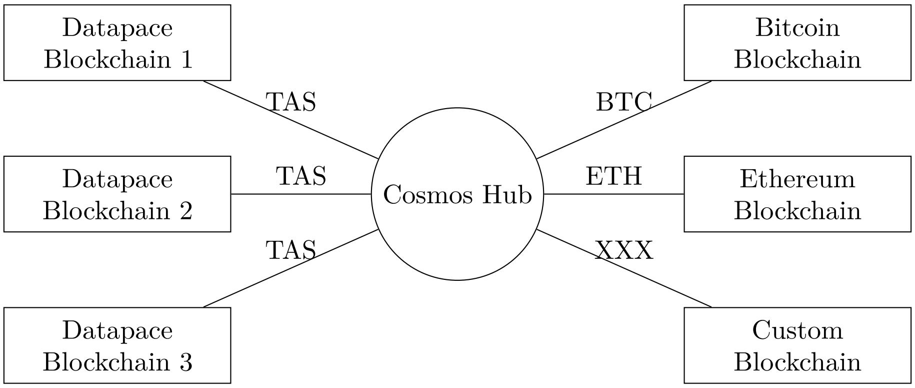

Figure 3: Datapace Sharding and Interoperability via Cosmos

details in a dedicated chapter.

- **Data integrity**: Leveraging on Datapace Hyperledger Fabric ERC-20 chincode that provides digital asset features, as well as native digital asset querying, Datapace implements mechanism that insures integrity of the data that flows through the system by taking it's digital fingerprint (cryptographic hash) and stores it in to the immutable blockchain database. This way system assures that critical data has not been tampered with. In the context of OTA firmware updates of safety-critical IoT devices or tamper-proof checking of already running software on such a systems (for example a braking system of a self-driving vehicule) this form of data security becomes quintessential.
- **Smart Contract**: Smart Contracts define a complex set of conditions under which data is exchanged. They are important part of Datapace system, and will be explained in detail in a dedicated chapter.
- **Network security (via PBFT consensus)**: In order to protect valuable digital assets and network infrastructure in the era of ever-increasing security threats[2], Datapace builds a decentralized network based on Byzantine fault-tolerant state and data replication algorithm. This way system can tolerate up to 1/3 malicious-acting nodes and assure network functioning under cyber-attack. Additionally, blockchain-structured data assure immutability and anti-tampering characteristics. Applying *Proof-of-Validated-Source* and *Proof-of-Stake* consensus, network is adding an additional layer of protection, incentivizing nodes to behave honestly and punishing badly behaving nodes. Based on these important features and technologies, Datapace builds high-security network that is capable to

[2] The number of records compromised grew a historic 566 percent in 2016 from 600 million to more than 4 billion. These leaked records include data cybercriminals have traditionally targeted like credit cards, passwords and personal health information, but IBM study[4] also shows a shift in cybercriminal strategies. In 2016, a number of significant breaches related to unstructured data such as email archives, business documents, intellectual property and source code were also compromised.

fully protect digital assets and insure secure protection of value exchanged through Datapace marketplace.

- **Auditing (via record immutability)**: Datapace enables monetary transactions, which are often subject to various regulations and can be examined by regulatory bodies. Thanks to the immutability feature of blockchain systems, Datapace system allows every organization participating in Datapace data market to have a proven track of records of all executed transactions.

To make system usable for the wide public, Datapace implements secure centralized wallet, similar to Coinbase[5]. Wallet, however, can be implemented also in decentralized fashion, as users can create accounts and transfer their funds to wallet of their choice.

In order to facilitate and standardize IoT data collection, Datapace system provides an IoT platform, based on Mainflux[3]. Mainflux IoT platform is integrated into Datapace system part called "Datapace IoT Platform", and exposes an API for sensor connection and management. Mainflux (and thus Moentasa IoT platform) equally provides IoT messaging and persistence capabilities, so all the data from sensors can be either offered in real-time or stored for historical usage.

Users of Datapace can use these interfaces to easily connect their sensors and thus enable data collection - this way they do not have to go through IoT system set-up, but can use Datapace platform as a service.

Additionally, once data enters the Datapace system via IoT platform, various types of processing and data analytics can be applied. Datapace will offer AI and ML algorithms to be applied to collected IoT data, so that users can enrich their data with with different type of intelligence prior to offering it to the market. This will boost the price of their data streams.

Finally, data coming through Datapace IoT platform is in standardized format - as per Mainflux specification, data is formatted in SenML[4] - a media type definition for representing simple sensor measurements and device parameters which comes in JSON and CBOR[5] flavors. This facilitates the operations for data consumers - they know up front what data format to expect and the same set of processing scripts, procedures and programs can be applied to various data streams.

It is important to note that use of Datapace IoT platform is not mandatory,

[3] Mainflux[6] (https://www.mainflux.com) is modern, scalable, secure open source and patent-free IoT cloud platform written in Go. It accepts user, device, and application connections over various network protocols (i.e. HTTP, MQTT, WebSocket, CoAP), thus making a seamless bridge between them. It is used as the IoT middleware for building complex IoT solutions.

[4] SenML[7] is a sensor markup language that aims to simplify gathering data from different devices across the network. It simply is JSON containing named events together with an associated value and unit.

[5] CBOR (Concise Binary Object Representation) is a binary data serialization format loosely based on JSON. It is defined in IETF RFC 7049[8]

even though it brings obvious benefits. Users can still obtain their own data via their legacy IoT installations and provide only access link to this data via Datapace market.

3.2 Technology Summary

Datapace integrates several open-source technologies which in combination provide a powerful system. An overview of technologies used is given in the table 1.

Table 1: Datapace technology summary

Hyperledger Fabric	Blockchain (Distributed Ledger). Provides PBFT consensus algorithm and P2P machine state replication. Adds security to Datapace blockchain. Immutability, querying, validator voting. Fast transactions.
ERC-20 Smart Contract	Provides ERC-20-compliant token as a form of divisible digital asset recorded by user-account balances.
Cosmos	Provides TAS token interchangeability. Provides Datapace blockchain scalability through sharding. Provides interoperability with other blockchain networks - like Ethereum or Bitcoin.
Mainflux	Provides IoT platform as a service. Enables IoT sensor and gateway connectivity and management. Provides IoT messaging, real-time and persisted data.

4 Data Verification

4.1 Overview

Datapace has unique-on-the-market solution for verifying the source of the IoT data. Based on the fact that Datapace and it's partners play one of the crucial roles in telecom equipment industry, an IoT gateways and edge computers were designed and connected with big number of sensors to serve as a verified and known IoT sensor data source.

Datapace installs these sensors in cooperation with network and telecom partners, or sends the certified equipment to various other partners for installation. Because these edge computers, IoT gateways and sensors contain known and certified hardware and firmware, often coupled with embedded GPS modules, system can be assured that data coming from these sensors is:

- Real-world data and not modified or generated “fake” data
- Coming from precise geographical location

Datapace partners that install and deploy this equipment will have an advantage on the marketplace, as their data sources will be marked as “trusted and verified”.

Moreover, since these partners made an economic investment and also entered in partnership with Datapace through various legal contractual agreements, they are allowed to run a validating node and participate in *Proof-of-Verified-Source* network consensus. Validators are rewarded for their work with TAS tokens.

4.2 Implementation

Datapace sensors are attached to Datapace gateways[6] and edge computers, or directly connected to the IoT platform.

Datapace provides IoT platform to monitor and manage IoT network and gather the data from the installed sensors. Once data is collected, it can be offered for sell by the user of Datapace system that has installed sensors (and/or gateways) and is the owner of the data.

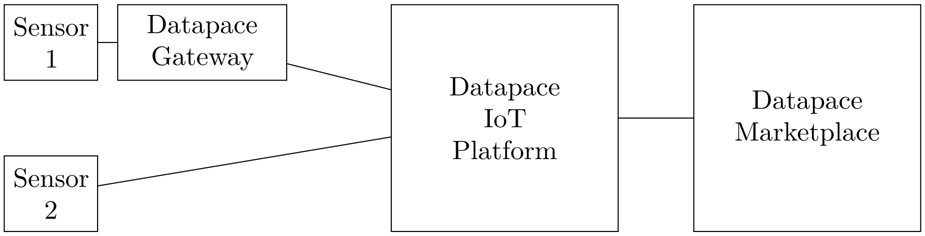

Figure 4: Datapace IoT device management via IoT platform

An additional benefit of enabling data connection via Datapace IoT platform is that users that choose this option can add various data processing and analytic services offered by Datapace system. Additionally, they can apply ML and AI insights to their data prior to offering them for sell, which would significantly augment the data price.

[6] Current Datapace IoT gateway implementation is based on novel EdgeX Foundry architecture. EdgeX Foundry is a vendor-neutral open source project building a common open framework for IoT edge computing. At the heart of the project is an interoperability framework hosted within a full hardware- and OS-agnostic reference software platform to enable an ecosystem of plug-and-play components that unifies the marketplace and accelerates the deployment of IoT solutions. More information can be obtained at project’s web address: https://www.edgexfoundry.org/.

5 TAS Token

TAS token is utility token of Datapace system. It is used to assure fair and secure functioning of the system, as well as to enable token economy on the Datapace data market.

Primary purpose of the token will be to fuel the system - it will be used to tokenize the value of digital assets (i.e. data) and facilitate their exchange. Equally, once the token economy is developed, TAS token will have a purpose in enabling the consensus mechanism based on *Proof-of-Stake*.

Data sellers will use TAS token as a representation of value of their digital data streams that are offered on the marketplace. Buyers will use TAS token to exchange it for selected data - they will transfer their TAS tokens to data sellers and obtain their digital assets in return.

6 Proof-of-Verified-Source and Proof-of-Stake

Datapace system employs two types of proof schemes that allow data providers and network participants to prove the data origin and quality as well as to enforce honest behavior of data sellers.

6.1 Proof-of-Verified-Source

In order to secure the network, Datapace system provides and original an unique approach called *Proof-of-Verified-Source*. This approach represents validator (miner) selection algorithm based on a proof of monetary investment in sensing hardware and networking equipment.

Due to the unique position on the market Datapace produces and delivers to the companies and network operators a specialized networking and sensing equipment - often an IoT edge gateway inter-connected with a lot of sensors. Since this hardware (and internal secure firmware) comes from known source (Datapace company), and since all equipment purchase and installation is done according to contractual agreements, everybody can be assured that this given data source is valid.

Because a company or an operator that purchases the equipment has to invest money, and also respect the written legal contracts, system can stay assured with high probability that they are incentivised to make fair decision (it is in their best of interest to keep the network secure and functional - otherwise their investment will be useless and they will suffer legal penalties).

Moreover, possibility to have *Verified Data Source* badge listed next to the data sources offered by these companies is an additional incentive for them to purchase the specialized sensors and other equipment.

6.2 Proof-of-Stake

Once TAS token economy is developed, a *Proof-of-Stake* consensus algorithm will be applied in order to additionally incentivise companies and individuals that run validator nodes and help secure the network.

Proof-of-Stake will equaly be used to enforce honest behavior of data sellers, because they will have to invest a monetary deposit (in form of TAS tokens). In case of malicious behavior (wrong data delivered, or data not delivered at all) deposit will be withdrawn by the system and bad actor will be punished.

7 Smart Contracts

A Smart Contract is a computer protocol intended to facilitate, verify, or enforce the negotiation or performance of a contract. The aim with Smart Contracts is to provide security that is superior to traditional contract law and to reduce other transaction costs associated with contracting.

Datapace platform provides possibility for users to define and deploy Smart Contracts that automate processes and formalize contractual agreements regarding various features of the system. One important feature, for example, is revenue sharing - every data seller can define a Smart Contract that will be signed by his partners and himself. Earnings obtained by selling this data stream will then be automatically divided between the parties, without further intervention from the seller and his partners.

Moreover, Smart Contracts enable fine-grained per-user and per-datastream conditions to be formalized. For example, new GDPR (*General Data Protection Regulation*)[9] laws by which by which the European Parliament, the Council of the European Union and the European Commission intend to strengthen and unify data protection of individuals, regulate the way that telecom operators or other companies can share user data. Since this data and it's sharing and monetization represent a core business of many companies (especially of those who's business model is based on advertising), a strict new relation between company and it's users is imposed and can be formalized and automated via Smart Contracts.

Datapace UI will enable defining these Smart Contracts in a simple manner though well-defined forms. Moreover, Datapace API will provide possibilities for these contracts to be defined and deployed programatically.

8 Data Integrity Through Anchoring

It is very well known feature of blockchains to offer immutable data storage. Once data is written in the blockchain it can not be changed (tampered with). This feature can be used to prove integrity of the data, which is especially important for OTA (*Over-the-Air*) firmware updates of safety-critical IoT devices or tamper-proof checking of already running software on various robots, machines, vehicles and similar.

In order to enable this feature as a service, Datapace implements an API on the top of its system that allows "anchoring" the data timestamp and cryptographic hash into the blockchain. This cryptographic hash essentially represents digital fingerprint of the data. Data hash can be recalculated and compared to immutable record in the blockchain at any later point, thus proving that the data has not be tampered with.

9 Future Work - Computing and Storage Tokenization

Besides data, a marketplace based on the blockchain can allow economy of at least two important resources:

- Storage
- Comuputing

9.1 Storage

Companies like **Storj**[10] or **Sia**[11] announced projects that strive to enable decentralized cloud. With low prices that would be a consequence of tokenized storage capacity offered by the various users in exchange of tokens, these companies can become a real competitors of SW giants in the cloud bussiness space - like Amazon or Google.

Datapace plans to integrate and maintain permissioned distributed file-system through wich Datapace users will be capable to offer and rent their storage space in exchange for TAS tokens.

9.2 Computing

Projects like **Golem**[12] or **SONM**[13] are working on decentralizing the computing power.

Based this ideas, but also on the ideas presented by **Blue Horizon** project from IBM [14], Datapace plans to enable Docker container based decentralized platform for deploying arbitrary software on the computing infrastructure offered and rented by Datapace users in exchange of TAS tokens.

10 Conclusion

Based on many reports[7], we can be sure of one thing: there is gold in the mountains of data. A way is needed to mine all this gold - a platform is needed to monetize all this data. Datapace is a an enabler that will unlock this huge potential.

Datapace builds decentralized marketplace based on blockchain, that is secure and scalable. It enables new token economy - TAS token will be used as an utility token of Datapace system, and will be used to enable fair and secure functioning of the system as well to enable trading facilities.

Datapace builds whole environment needed for quick adoption of the system: UI, wallet, API and SDKs. This will lower adoption barriers and lead to the higher popularity of the system, which will in turn incentivise the economy based on TAS token.

Due to unique positioning, Datapace provides specialized senor hardware, and employing various patent-pending techniques assures that data sources are verified. Moreover, through specific AI and machine learning algorithms, Datapace system assures that all data streams can be unified in format and prepared for easy consummation. This brings clear advantage of Datapace comparing to all existing competition.

Datapace will be go-to marketplace for data monetization - any data, anywhere.

11 Contact

Website: https://www.datapace.io

E-mail: info@datapace.io

11.1 Social Networks

Twitter: @DatapaceMarket

[7] IDC says that worldwide revenues for big data and business analytics will grow from $130.1 billion in 2016 to more than $203 billion in 2020, at a compound annual growth rate (CAGR) of 11.7%[15]. In addition to being the industry with the largest investment in big data and business analytics solutions (nearly $17 billion in 2016), banking will see the fastest spending growth. New report from McKinsey & Company's Global Institute is trying to put a real dollar amount to the global IoT market. In the report's estimation, IoT has the potential to be worth between $3.9 and $11.1 trillion by 2025[16].

LinkedIn: https://www.linkedin.com/company/datapace/

Facebook: https://www.facebook.com/datapace

Acknowledgments

This work is the cumulative effort of multiple individuals within the Datapace team, and would not have been possible without the help, comments, and review of the collaborators and advisors of Datapace. Drasko Draskovic ad George Saleh published the original Datapace ideas in 2016, laying the groundwork for this work. Special thanks goes to the community contributors: Maxime Aubanel, Aleksandar Likic and Marta Piekarska from IBM's Hyperledger team - thanks for hanging with us on IRC channel, guiding us through Fabric internals; Jae Kwon and Ethan Buchman, Tendermint and Cosmos creators - thanks for moving blockchain tech forward; we appreciate huge help and contributions given by Janko Isidorovic and Mainflux team - this magnificent piece of IoT technology became one of the essential architectural blocks of Datapace system. We also thank all of our collaborators and advisors for useful conversations.

References

[1] IBM Marketing, "10 Key Marketing Trends for 2017." 2017 [Online]. Available: https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN

[2] J. Kwon and E. Buchman, "Cosmos - Internet of Blockchains." [Online]. Available: https://cosmos.network/

[3] Wikipedia - The Free Encyclopedia, "Shard (database architecture)." [Online]. Available: https://en.wikipedia.org/wiki/Shard_(database_architecture)

[4] IBM X-Force, "IBM X-Force Threat Intelligence Index." 2017 [Online]. Available: http://www-03.ibm.com/press/us/en/pressrelease/51946.wss

[5] "Coinbase." [Online]. Available: https://www.coinbase.com

[6] D. Draskovic, J. Isidorovic, D. Mijic, and N. Marcetic, "Mainflux IoT Platform." 2015 [Online]. Available: https://www.mainflux.com

[7] C. Jennings and Z. Shelby, "Media Types for Sensor Measurement Lists (SenML)." 2017 [Online]. Available: https://tools.ietf.org/html/draft-ietf-core-

senml-12

[8] C. Bormann and P. Hoffman, “Concise Binary Object Representation (CBOR).” 2017 [Online]. Available: https://tools.ietf.org/html/rfc7049

[9] Wikipedia - The Free Encyclopedia, “General Data Protection Regulation.” [Online]. Available: https://en.wikipedia.org/wiki/General_Data_Protection_Regulation

[10] S. Wilkinson, “Storj - A Peer-to-Peer Cloud Storage Network.” 2016 [Online]. Available: https://storj.io/

[11] D. Vorick and L. Champine, “Sia: Simple Decentralized Storage.” 2014 [Online]. Available: https://sia.tech/

[12] “Golem.” [Online]. Available: https://golem.network/

[13] SONM Team, “SONM - Supercomputer Organized By Network Mining.” 2017 [Online]. Available: https://sonm.io/

[14] IBM, “Blue Horizon.” [Online]. Available: https://bluehorizon.network

[15] IDC, “Worldwide Semiannual Big Data and Analytics Spending Guide.” 2015 [Online]. Available: https://www.idc.com/getdoc.jsp?containerId=prUS41826116

[16] McKinsey Global Institute, “The Innternet of Things - Mapping The Value Beyond The Hype.” 2015 [Online]. Available: https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/the-internet-of-things-the-value-of-digitizing-the-physical-world