

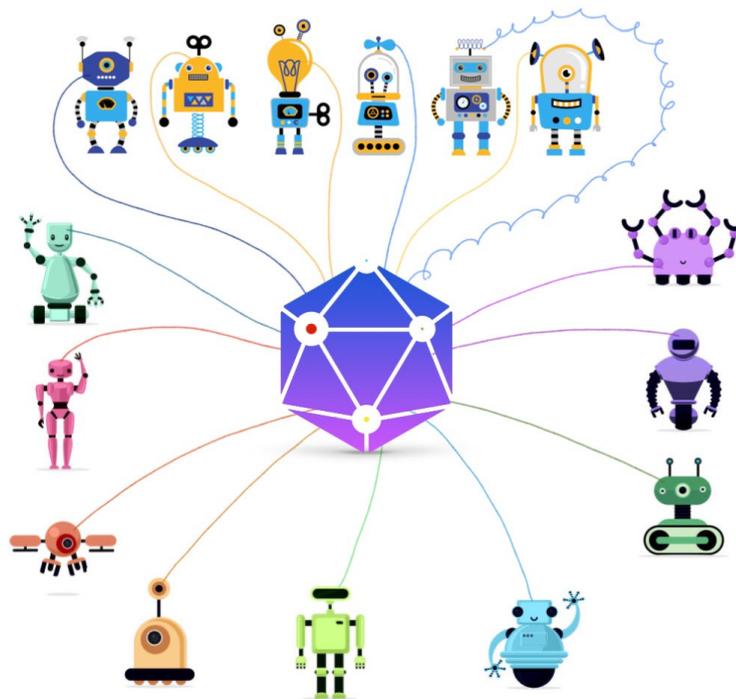
DATALAC.COM

Connecting Dataverses

founders@datalac.com

Abstract: This document describes Datalac.com - a community-powered platform for creating and developing large Datalakes from public data sources on social media based on the \$DTVN token blockchain. Each Datalake created through datalac.com can help form a knowledge system (we call a Dataverse). The vision of datalac.com is not only to help build knowledge systems for global individuals and businesses with data needs but also to form a foundation of connecting those knowledge systems. Datalac.com helps to

easily accumulate and share valuable knowledge, helpful information and interesting data. Developing with the fashion of web 3.0, Datalac.com also helps to quantify the value of data and knowledge. Datalac's Data To Earn game model provides individuals who regularly use the internet a sustainable, effective method of earning additional income and provides organizations and businesses with new sources of valuable and updated data.



1. Data is the new water

“Data is the new oil”. This quote has only appeared since 2006 but quickly received the approval of many people online. Along with the development of mobile internet and smart-phones, data flows are constantly born and circulated everywhere on the internet, affecting the lives of netizens in many different ways. Big Data technologies combined with Artificial Intelligence also form new generation software engines¹, fueling the blossoming of the Fourth Industrial Revolution². Data has become one of the most important resources, helping to shape the present and future of modern societies.

The importance of data has long been recognized by businesses in developed countries. Strong technology enterprises in developed countries have collected and accumulated data for many years through global software, forming large Datalakes³ as resources for testing and implementing many pioneer and breakthrough business models on the internet. This process inadvertently creates a "data drought" in many places, because internet users concentrate mostly on popular software, making the application of data mining technology in less developed countries becomes difficult.

Not just like oil, data is similar to water resources. Water is an important resource for all societies. During the Second Industrial Revolution, water was stored in large reservoirs, concentrated energy and from there creating electrics, coordinated to power many types of electric motors for production. The 4V (Volume, Velocity, Variety, Veracity) characteristics of Big Data are also very close to the characteristics of a water stream. And data accumulation and collection offer just as many benefits and resources as doing with water resources. The data of our time is therefore also stored by organizations and businesses in Datalakes, creating knowledge systems to develop powerful data-driven and artificial intelligence solutions.

¹ Software 2.0

<https://karpathy.medium.com/software-2-0-a64152b37c35>

² The Fourth Industrial Revolution

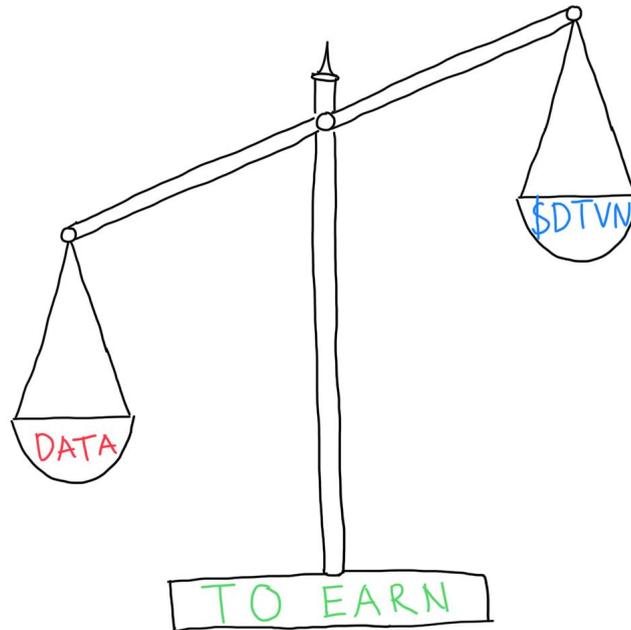
<https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/>

³ Google Cloud launches BigLake

<https://techcrunch.com/2022/04/05/google-cloud-launches-biglake-a-new-cross-platform-data-storage-engine/>

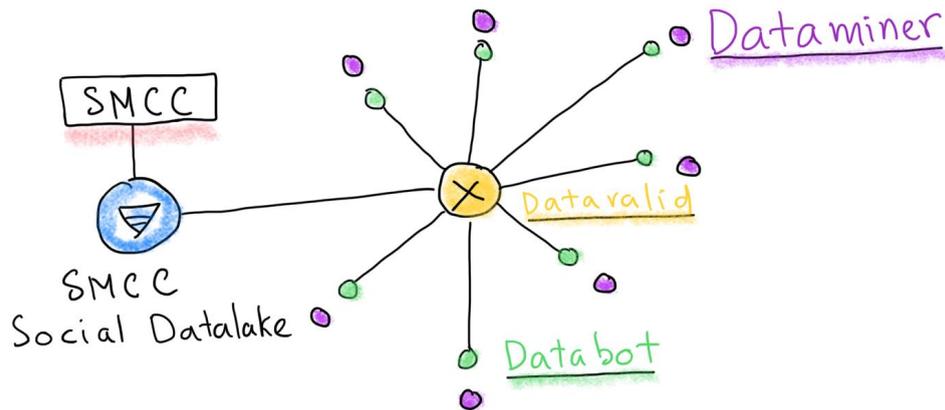
The mission of datalac.com is to create a network of Datalakes by the power of the community, based on blockchain technology, following the Data To Earn model, serving the global needs of big data. Lac in French means lake, the name DATALAC also implies that many Datalakes will be built, forming different but all connected Dataverses globally.

2. The “Data To Earn” Game



To form big Datalakes, DATALAC develops the Data To Earn game model with the following main tasks:

- Developing the Databot software robot, which collects public data on public social media channels under the control of Dataminers.
- Developing Datalake software to help store, analyze and share big data, build Knowledge Systems based on data.
- Developing Datavalid software to help create an automatic system for assigning tasks, synthesizing - evaluating results and making decisions about rewards for Databot operators.
- Creating \$DTVN token as a tool to quantify the value of collected data and transparently distribute benefits to those involved in operating Databot and those who need to build Datalake.
- Build a community of Dataminers who operate Databots.



Initial cycle of Datalac with only Datavalid Zero

When more organizations and businesses need to join Datalac.com, new Datavalid and Datalake servers can be installed. The newly established Datavalid allows to receive registrations from Dataminers and register with existing Datavalid to conduct the exchange of data collection requests and rewards in \$DTVN tokens with each other according to the "Data To Earn" mechanism.

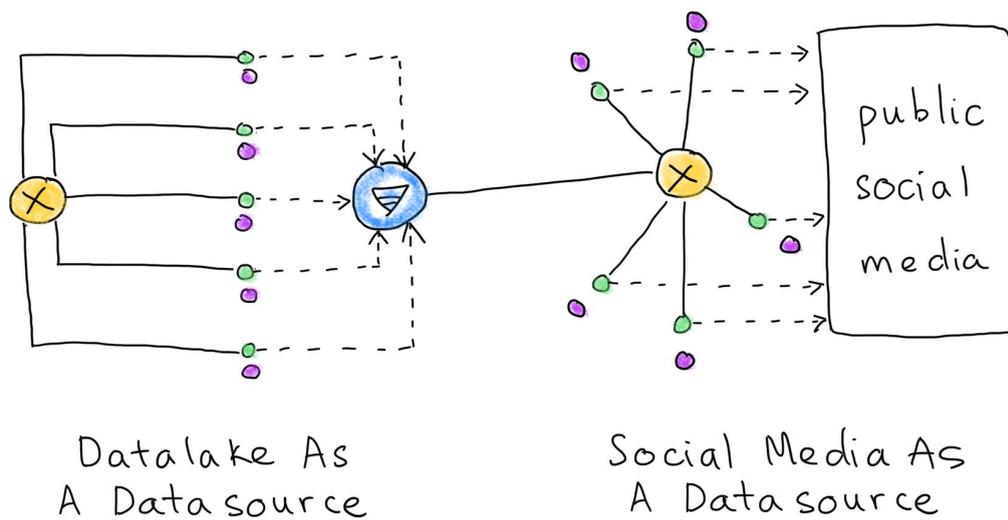
Further, Dataminer can set up its own Personal Datalake to meet the data mining needs of individuals also under the "data to earn" mechanism. In the following sections of the content, we would like to describe in more detail the basic components of the Datalac.com system.

4. Dataminer

Dataminers are users who join Datalac's "data to earn" program, directly "raising" Databot to do distributed data collection tasks. Each user registers with DATALAC and validates his email address, blockchain wallet address to start participating in the program. If you imagine "data to earn" as a game, then Dataminer is an important group of players.

After registration, Dataminer installs Databot software on your computer or phone and log in to initially participate in "data to earn". Each Dataminer can install more than 1 Databot if needed. In order to encourage and develop the Dataminer community, at the beginning of the project, Datalac did not charge the operation of the basic Databots by providing the Databot NFT level 1 for free to the Dataminers under the Free To Play policy.

Individuals, organizations and businesses that install and operate Datalake themselves can also maintain their own Dataminer team through robots that navigate the data source to Datalake. In other words, Dataminer that owns Datalake can allow its Databot to get saved data straight from Datalake to meet demand instead of having to read data on the internet.



Two type of datasources for a Databot: from a Datalake or from the Internet

5. Databot và Databot NFT

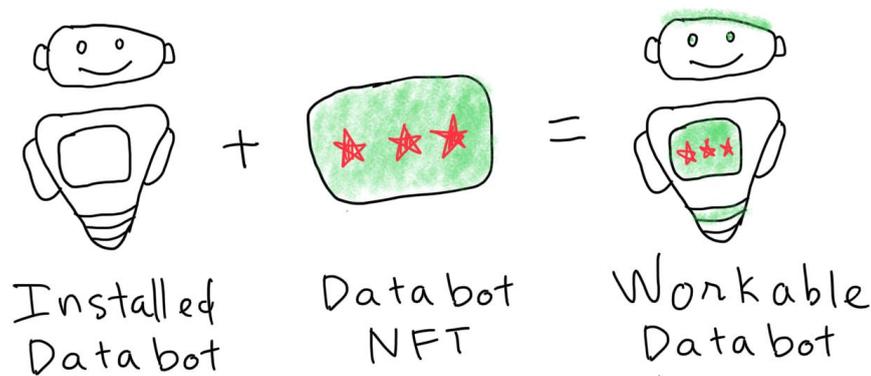
Databot is the software responsible for collecting distributed data in the Datalac network. Databot is designed as a compact software, works on computers, phones with limited configuration, on many different operating systems. When enabled, the Databot connects to the pre-registered Datavalids, reads the public content requested by the Datavalids, acquires the data, and returns the results.

Each Databot is represented by an NFT allocated by Datalac. The Databot NFT contains information about the level of the Databot that the NFT represents. In order for the Databot to receive a data collection request from a Datavalid, the Dataminer needs to register the Databot NFT with the Datavalid via the NFT's ID.

On each Datavalid, each Databot NFT represents only 1 Databot. Dataminer can own multiple NFT Databots. Databot NFT can advance to higher levels by pooling multiple Databot NFTs together. Databot NFT is subject to wear and tear and can be repaired and exchanged at Datalac.com Online Store.

The Free To Play policy is applied by Datalac to encourage Dataminers to participate when the new project is established to provide Databot NFTs for free. However, these NFT Databots cannot be traded and upgraded. For higher-level NFTs, Dataminer needs to go to the Datalac Online Store.

When completing a data collection task, if the data from Databot submitted to Datavalid meets the conditions, the Dataminer that owns the Databot is rewarded with an amount of \$DTVN corresponding to the level of the Databot NFT. The higher the level of the Databot NFT compared to the Databot NFT of other Databots sending the same data sample, the more \$DTVN will be received by the Dataminer.



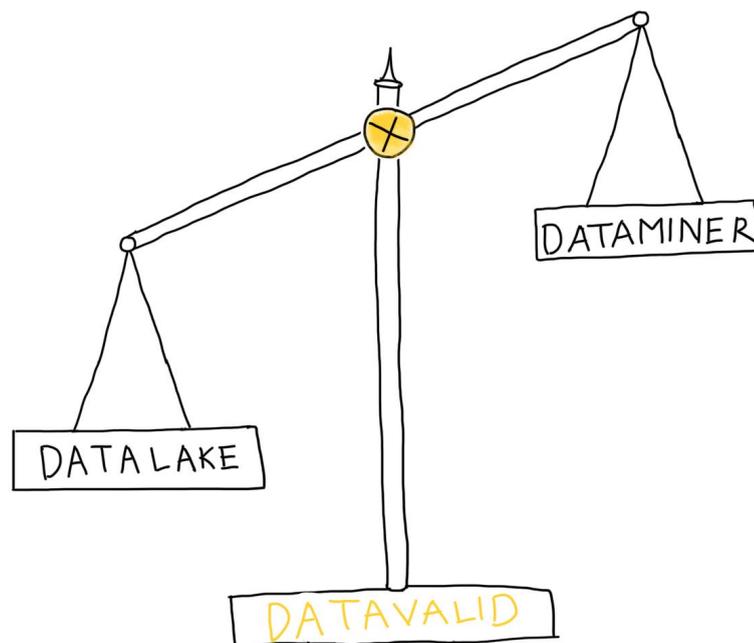
Databot must have Databot NFT as a permission to play

Databot has many different versions for the social media platforms to collect data. Databots can also directly connect to Datalakes allowing them to collect data, thus opening up the possibility of easy data exchange between Datalakes through the coordination of

Datavalids. Different versions of Databot can use the same Databot NFT. Databot's source code is open so that the community can jointly edit and develop new features with the Datalac project.

6. Datavalid

Datavalid software is the system that plays a central role in the Datalac.com network. Datavalid is responsible for connecting data supply and demand in Datalac, allowing the distribution of data collection requests to Databots; verify the quality of the data sent; calculate the amount of reward tokens for Dataminers per task; classifying Databot and Dataminer; share, respond data with other Datavalid; provides aggregated data to the Datalake.



Datavalid works on servers of individuals, organizations and businesses wishing to collect, store and enrich data. Individuals, organizations and businesses with this need first need to store a large enough amount of \$DTVN to pay rewards when receiving data and to ensure trust for Dataminers. At the same time, Datavalid needs to maintain a list of addresses that need updating along with the frequency of updates. Datavalid receives service subscriptions from Databots fed by Dataminer. The Databot NFT of each Databot is validated against

ownership on the Dataminer blockchain wallet. After successful registration and connection, Databot is ready to receive Datavalid's requests to scan data.

To ensure the quality of received data, Datavalid applies the mechanism of group data collection and time frame cross-validation. For each address that needs to update data, Datavalid sends a random group of Databots belonging to many different Dataminers and possibly other Datavalids associated with them. The fastest response data after a time frame t are included in the review list. During the review process, data from different Databots and Datavalids are cross-matched to select only the same data. Databots that respond to data that are radically different from the majority will not be paid \$DTVN and marked for spam detection. Databots that respond with duplicate data are included in the \$DTVN payout list.

\$DTVN rewards each successful Databot at each calculated data task depending on the Databot NFT level of that Databot in percentage correlation with the Databot NFT level of other Databots. The higher the level of the Databot NFT, the better the dataminer owning the Databot has the chance to receive more \$DTVN, the weights of the ranks follow the distribution of the Fibonacci sequence.

$$D_{bx} = \frac{r(b_x) D_{\Sigma}}{\sum_{i=0}^n r(b_i)} ; x \in [0, n]$$

The formula for calculating the amount of \$DTVN received by each Databot for completing a data task, passes a cross-check. D_{Σ} is the total amount of \$DTVN for all n Databots completing the task; bx is the x th Databot in the group; $r(bx)$ is a weight function that depends on the rank of Databot NFT fitted to Databot bx , $r(bx)$ distributed according to the Fibonacci sequence

The results of each data collection task on Datavalid are authentic data records, which are transferred to Datalake systems associated with Datavalid for storage and exploitation. Datavalid of the Datalac.com project is also developed as an opensource project to make it easier for the community to participate in development and installation when there is a need to deploy.

7. Datalake

Data, when updated continuously, will form a flow. If we imagine “Data is the new water”, we can compare data flows to streams and rivers. To store energy from water resources, people build large dams to store water. Large water reservoirs help the material in the water to naturally deposit mud and sand, creating clean water to run generators, forming hydroelectric power plants. Hydroelectric reservoir is one of the central works of the Second Industrial Revolution, providing stable electricity for electric motors operating in the factory.

Datalake also has the same positive features as Hydroelectric Lake, storing big data from data flows. Datalake has the ability to optimize, automatically classify the types of data received based on the network graph architecture. In Datalake, data points representing any entity are arranged in a structure that connects to the data points of the related entity. Information coding models using Artificial Intelligence implemented in Datalake allow automatic classification and clustering of data, creating measurements to bring data points with high similarity close to each other. Thus, when it comes to mining, individuals, organizations and businesses can easily find blocks of high-value data without much effort in filtering and labeling data.

Datalake also provides high-speed information retrieval. Data is stored centrally but divided into many pieces and many copies to ensure integrity and speed of access. As a result, Datalake not only allows exploiting and finding valuable knowledge, but also allows data communication with other Datalakes, helping organizations and businesses easily enrich their existing Datalake.

Datalake of DATALAC project is inherited from Social Datalake of SMCC.VN - the pioneer social listening software system in Vietnam, the product won the first prize of Vietnam Talent Award 2016. SMCC was established in 2011 and has been applied to many national projects. Under the cooperation with Datalac, SMCC Social Datalake has been upgraded and transformed into an opensource project so that individuals, organizations and units wishing to participate in Datalac.com can easily install a data receiving, processing and visualizing infrastructure.

Depending on the volume of data to be stored and mined, Datalakes may require different sized hardware infrastructures. At a minimum, Datalake needs to operate on only 1 standalone server. At a high level, Datalake may require a multi-server cluster, with separate components for storage, search, training, and Artificial Intelligence. When large organizations and enterprises have a need to deploy Datalake, Datalac.com can also provide the necessary advice and support.

Not only targeting Datalake at the organizational and enterprise level, the Datalac.com project also aims to develop Datalake specialized for individuals to collect, store, enrich, benefit and inherit knowledge from their personal data.

SMCC Social Datalake - First Datalake of the network

To initiate, the Datalac.com project cooperates with SMCC.VN Data Lake to transfer the data collected by the Dataminer community into SMCC Datalake. SMCC.VN is a software system for analyzing social media data developed by InfoRe Technology from 2011 to present. Operating as a software as a service (SaaS), SMCC specializes in scanning news websites, online newspapers and public pages on Facebook in Vietnam to aggregate daily information; The aggregated information is processed by Artificial Intelligence, forming a high-speed statistical search engine to serve the information retrieval needs of data users to protect the brand, detect media crisis, social trends, customer care... and especially searching and optimizing customer files for marketing campaigns of businesses.

In 2016, SMCC won the First Prize of the Vietnamese Talent Award and now has more than 28,000 users, accessing statistical information about most areas of society every day on billions of data records. Celebrating its 10th anniversary, SMCC is aiming to go overseas to

provide millions of data analysts with modern data-driven tools and ways of working. SMCC owns an automatic robot system capable of scanning information on a large scale, with a daily capacity of up to tens of millions of records, operating continuously 24/24. However, due to the explosion of mobile internet, the amount of information on the internet is increasing, SMCC's automatic robot system only covers information about the Vietnamese market. In order to serve the markets of countries in the region and many other countries around the world, SMCC needs a more powerful information collection robot system.

DATALAC and SMCC are two necessary pieces of the puzzle. On the one hand, SMCC makes the data collected in the early stages of the DATALAC project meaningful, valuable, and immediately exploitable. On the other hand, DATALAC helps attract many participants to collaborate with SMCC to scan data from many different markets and benefit, so that SMCC can develop to many major markets around the world.

8. Personal Datalake

Personal Datalake is the software supports Dataminers to accumulate and perform, exploit personal data without going through the \$DTVN token thanks to being directly connected to the Databots. Personal Datalake enhances benefits for Dataminers in the digital age. Details of individual Datalake will be updated by the project in the near future.

9. Data To Earn Gameplay

- DATALAC's "Data To Earn" model can be imagined as a game between the community of people who are interested in public data on social media according to the main rules of the game as follows:
- Participants exchange in-game demand types with two main Datalac tokens: \$DTVN and Databot NFT.
- How to play: Using tokens as an intermediary to meet the supply and demand of data, between those who can update data regularly (Dataminer) and those who need to update data regularly to serve data mining (Datalake implementers).
- Dataminer purchases Databot NFT from DATALAC and installs and operates Databot.

- The Databot registered with a Datavalid will receive a request to collect data from that Datavalid (via the Databot NFT ID).
- Each request from Datavalid is randomly sent to n Databots that are ready to go. Databot conducts automatic data collection on request and sends feedback results to Datavalid
- The earliest m Databot that responds before timeout t will be rewarded with \$DTVN tokens if the data sent by m Databot pass the cross-validation.
- The transaction history of each Databot is stored and used by Datavalid as a basis for rewards, punishments, and Databot spam filtering.
- Each Datavalid determines its own reward price, which can be based on the basic reward price that the \$DTVN fund for Datalake owns Datalac project: The total fund accounts for 50% of the project's tokens and half of the reward value every new data cycle.
- \$DTVN rewards are divided among m Databots at each data collection task depending on the level of each Databot NFT in group m .
- The level of each Databot NFT can be upgraded based on merging with other Databot NFTs.
- When merged to form a Databot NFT with a higher rank, the component Databot NFT will be destroyed.
- Databot NFT can be traded and exchanged on the exchange system between Datalac project participants. At any Datavalid, each Databot NFT can only equip 1 Databot at a time.
- Databot can collect data by reading public information on social media or reading accumulated information directly from Datalakes (if read permission is granted).
- Data from Datavalids is transferred to Datalakes to serve data mining operations.
- The output parameters of the Databots under the management of each Dataminer can be monitored on Datalac's mobile app and Datalake's personal mobile web.
- The source code of the technology components in Datalac is opened so that individuals, organizations and businesses can easily join the Datalac network, proactively install their own Datalake and Datavalid when needed.

10. Tokenomics

The Game “Data To Earn”

Data To Earn is the core game of the Datalac project. Data To Earn is designed as a game between those who need to update data and those who need to collect \$DTVN tokens. The Datalac project is expected to issue 1 billion \$DTVN tokens as an intermediary tool for meeting supply and demand in the game "data to earn".

The dynamic of "data to earn" is the balance of supply and demand between the supply side who are able to contribute to data scanning (Dataminers) and the demand side who need to set up, Build and maintain Datalake. The supply side maintains the Databots to supply the data to the demand side. The demand side pays \$DTVN packages through a smart contract on Datavalid in exchange for data. Datalake owners themselves can also turn their Datalake into a data source for other Datalakes (if Databots are allowed to connected). \$DTVN is the central peer-to-peer tool of the Datalac project, helping to meet the demand and supply of data of all parties and thereby forming tokenomics around the community of Datalac platform participants.

Thanks to Data To Earn, Datalac can be visualized as a platform that meets data supply and demand in a distributed form. To start forming a tokenomic around Datalac, the intermediate parity tool \$DTVN is not enough, due to the "chicken - egg condition" of supply and demand. When the number of people wishing to build Datalake is not large enough, the source of payment of \$DTVN is limited, making it difficult for the number of Dataminers participating in the project to provide data to grow. The low number of Dataminers makes the project's ability to provide data limited and difficult to attract those who want to build Datalake to participate in the project.

To solve this problem, Datalac uses 50% of the \$DTVN token (equivalent to 500 millions \$DTVN) to regularly pay Dataminer rewards on the first Datavalid (Datavalid Zero). This amount of \$DTVN is split in half at the beginning of each cycle, half is used to reward Dataminers in the current cycle, half is used for the next cycle according to the principle of halving. Each cycle is defined by the total number of correct records that Datavalid Zero has

verified. The number of records in the first cycle is 10 billion records, in the subsequent cycles, the number of records is equal to the sum of the previous 2 cycles.

| Cycle | Verified Data Records on Datavalid 0 | Exchanged \$DTVN on Datavalid 0 |
|--------------|---|--|
| 1 | 10 billions | 250 millions |
| 2 | 10 billions | 125 millions |
| 3 | 20 billions | 62.5 millions |
| 4 | 30 billions | 31.25 millions |
| 5 | 50 billions | 15.625 millions |
| 6 | 80 billions | 7.8125 millions |
| ... | ... | ... |

The periodic half-life payout on Datavalid Zero helps to solve the following problems:

First, create an initialization mechanism for the Datalac project at the beginning, avoiding the chicken-egg condition. When there is no Datalake of active users, the Dataminer community participating in Datalac still has a clear and beneficial operating goal: participate in collecting 10 billion data records in exchange for 250 million \$DTVN tokens.

Second, create development needs for the Dataminer community at the first time. Being an early Dataminer with the Datalac project helps Dataminer to benefit from a higher exchange rate per data record for \$DTVN in later cycles. The sooner the Dataminer joins the better. Having multiple Dataminers makes it easy to meet the needs of Datavalids and new Datalakes.

Third, at each cycle, the payout rate on Datavalid Zero helps to provide a reference exchange rate for the entire Data To Earn market. Datavalids that pay higher than the rate of Datavalid Zero may attract more Dataminers to participate but may be less attractive to those who need to build Datalake.

Other token funds of the project

The Dataverse project's \$DTVN token has a total supply of 1 billion units, distributed into funds as follows:

- Data To Earn Fund: 50%
- Founding Team Fund: 12%
- Angel Sponsor: 6%
- Seed Funds: 2%
- Strategic Cooperation Fund: 5%
- Direct Cooperation Fund: 5%
- IDO / IEO IDO Fund: 5%
- Marketing Fund: 10%
- Reserved Fund: 5%
- Treasury Fund: 0%

Treasury Fund is a fund formed when operating the project, formed from the sale of Databot NFT.

Token payment schedule

Accompanying this document, we have a detailed estimate of the token payment schedule and token vesting plan for the Datalac project on each fund. For a detailed estimate, please contact the founders team at founders@datalac.com

11. Development Plan

Planned milestones from 6/2022 to 3/2023

June 2022

Design tokenomics, vesting table, pitch deck

Build whitepaper version 3 of Datalac

Build game play for Datalac

Recruiting and forming a specialized marketing team for Datalac

Recruiting and forming a team of Datalac development technicians

Overall technical architecture design

Deploy data acquisition infrastructure

July 2022

Start building the Delivery System - receive data scanning jobs with Data Picker

Start building the System of scoring, ranking users and redeeming reward points

Completing the document system for communication about the Datalac project

Develop an overall marketing plan for the Datalac project

August 2022

Experimenting with Data Picker - receive data scanning jobs with Data Picker alpha version

Testing the Pointing System, User Rating and Redeeming Alpha Version Rewards

Design and build Databot NFT . system

Build a testnet to test the allocation of DTVN, Databot

Finding and contacting investors for the Strategic & Private Sale round

September 2022

Forming a community of Datalac participants from testers

Completing the Data Picker - receive job scanning system with Data Picker

Completing the System of Scoring, User Rating and Redeem Points

Build a payment system, buy a robot farming account with tokens

Implement a marketing campaign to expand the tester community

October 2022

Start building the Databot trading and management system

Developing mobile apps to serve players to manage data to earn

Upgrading infrastructure for data acquisition

Closing the list of Strategic & Private Sale partners for the first round

November 2022

Internal issuance of DTVN tokens and early Databots
Completing the review process, redeeming reward points
Testing mobile apps to serve players
Start airdrop to pay DTVN tokens to testers and transfer to founders, early investors
Testing Databot purchasing and management system alpha version
Make the first round of Strategic & Private Sale with at least 1 partner

December 2022

Complete the mobile app, put it on the store
Completely replace tokens on testnet with DTVN and Databot
Smart contract programming and blockchain technical solutions
Deploy digital marketing campaigns to the blockchain community
Looking for partners CEX & DEX to cooperate to completeIDO / IEO / INO procedures.
Start collecting list of IDO / IEO / INO subscribers.

January 2023

Working with blockchain marketing networks, launching partners to promote the Datalac project
Maintain digital marketing campaigns about the project
Testing and perfecting technical systems
Deploy a security solution and manage token wallets, divide tokens into different funds

February 2023

Open and sell accounts for raising robots with DTVN
Edit and optimize the technical system
Promote marketing for IDO / IEO in March 2023
Achieved the expected number of IDO/IEO subscribers
Cooperate with partners to organize events, events for media

March 2023

IDO / IEO project Datalac (distribute DTVN and Databot)

Turn the Datalac project into a community project and officially operate.

12. IDO / IEO

Datalac.com project aims to distribute tokens to the community by March 2023. This timeline may fluctuate based on the market and the development of the Datalac project in the coming time. We will regularly update this whitepaper to keep you informed about Datalac's extensive token distribution schedule.

13. Future Orientation

Vision for the future of Datalac is a world where data plays a pivotal role in people's lives. Powerful data mining and profiting technologies such as Blockchain and Artificial Intelligence will continuously evolve and transform lives. The need to use data in society will become a broad-spectrum need, like today's demand for electricity, water, and internet. In that world, individuals and businesses need to be equipped with tools to easily store data resources, select information from the stored data; extract, visualize knowledge and transform discoverable knowledge into easily exchangeable assets. Genetic data, social relationships, consciousness and physical condition, etc. will be precious data that each individual needs to own, master, manage and exploit. Social, market, and customer data are also valuable resources for every business, helping businesses develop and innovate in business. Datalac.com wants to participate in this process and not only wants to create a market for public data sources, but also wants to create good tools for data mining, benefiting individuals and businesses .

14. Stay Update

This is a document that is regularly updated by the founders of Datalac. If you would like to keep an eye on this document, you can sign up for the Datalac Maillist at <https://datalac.com/maillist> . We will send you periodic notifications about new information on the project.