

Khamseen Network v1.0: an ecosystem for a new digital economy

Khamseen Network Team

contact@khamseennetwork.org

October 1st 2022

Contents

1	Introduction	2
2	Khamseen Network: a proposition for a more balanced data economy	5
2.1	Search engine, private data and data trusts	5
2.2	Data marketplace	7
2.3	Why another token?	7
3	Technical aspect of Khamseen Network Ecosystem v1.0	8
3.1	Khamseen Network Data Platform	8
3.1.1	Data trust modules	8
3.1.2	Hybrid search engine	9
3.1.3	Digital advertising	10
3.1.4	Data marketplace	10
3.1.5	Data processing task requests	13
3.1.6	Contributors and compensation scheme	13
3.2	Khamseen Token	14
3.2.1	Token summary	14
3.2.2	Minting mechanism and compound effect	15
3.2.3	Token crowdsale	16
4	Roadmap	16
5	References	17
6	Important Information	19

Abstract

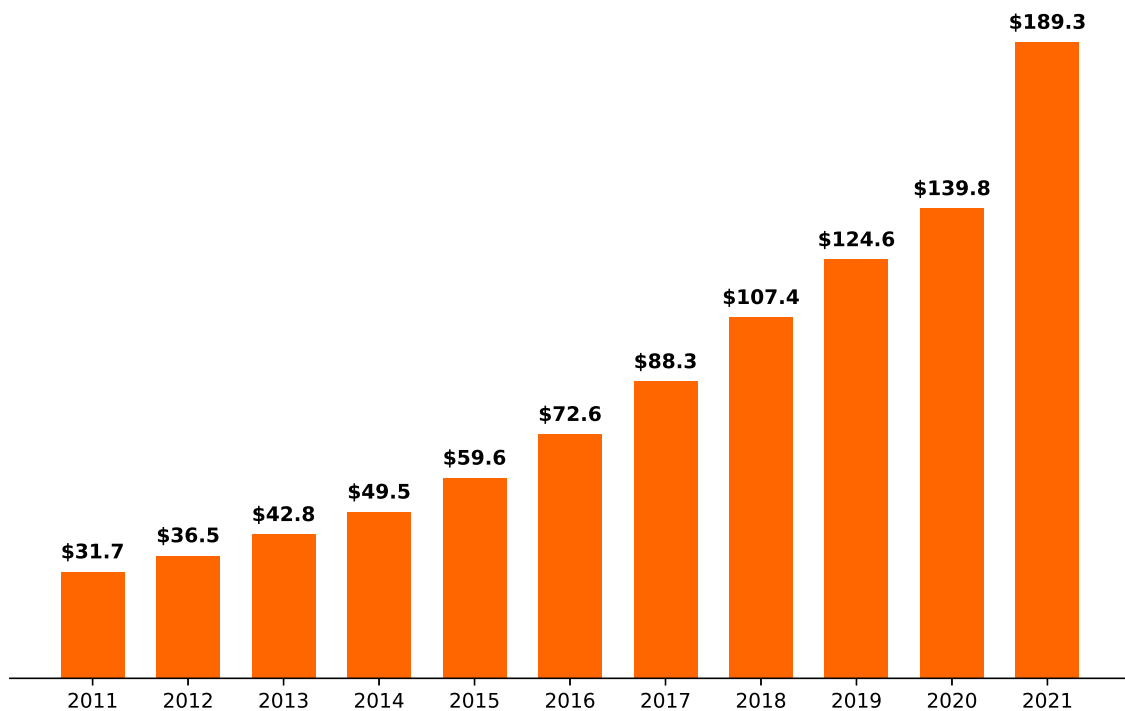
With increasing and more ubiquitous integration of digital technologies into everyday life, personal data has become both a huge source of revenue and an increasingly more controversial aspect of the tech industry due to its sensitive nature. After multiple controversies and clear failures from tech companies regarding this issue, governments and startups are both tempting to remedy the power-asymmetry between data subjects and data controllers as well as user alienation from their own data. In this paper, we present the Khamseen Network and the first phase of our ecosystem. We propose a new model for the data economy putting focus on data dignity and user empowerment. We will leverage both centralized and decentralized technologies to create a hybrid ecosystem to enforce transparency and fairness to the data economy both for companies and users. The first phase of this ecosystem will be based on four applications linked to each other: a library of data trust modules helping users pooling and monetizing their data, a data marketplace for transparent and secure transactions, a more empowering search engine tapping into both Web2 and Web3 contents and finally a new token designed to be a more meritocratic store of value as well as an optional utility token for our ecosystem.

1 Introduction

Some of the largest internet companies existing to this day have been building their business model on pervasive digital ecosystems based on search engines, email services, social media and mobile applications yielding a massive amount of personal and private data from their users. For years, and as the world continued to digitalize, those companies have been able to handle and monetize that data with virtually no oversight and adequate restrictions. This led to two major consequences. The first consequence was the rise of a digital economy based on data as both a commodity to be sold and a crucial tool for the growth of a company. This translated to continuous effort and massive investments from those companies to build expertise and infrastructure to better manage, analyze and monetize the data generated from their own ecosystems. The second consequence of this trend was the realization of the sensitive nature of private data and the importance of implementing data privacy regimes. Users are put in an increasingly vulnerable situation where their online self-identities are being monetized with no compensation and their data are being processed with higher levels of granularity and complexity.

This new "data market" represents such a massive source of revenue from tech companies that it has been valued at \$75 billion in Europe and at \$185 billion in the US in 2019 while its measurable impact on the economy i.e. what we can call the "data economy" has reached \$400 billion in the same year for Europe alone according to a recent study conducted by the European Commission [7]. According to this same study, projections for 2025 estimate that the data market and the "data economy" in Europe alone would reach up to \$107 and \$827 billion respectively. According to a McKinsey study [20], data economy projection for 2025 would even reach between \$845 billion and \$2.5 trillion globally.

Data monetization can take two different forms: direct and indirect [26]. Direct data monetization consists in selling or bartering data to external third-parties while indirect data monetization consists of improving already existing business processes or building new ones with insights gained from this data [19]. One third of companies which monetize data do it in a direct way [23]. Those data transactions are usually made on data exchanges or data marketplaces but can also be made in more secretive deals, between close partners in related industries through pools of shared data [17, 21], which can raise even more the question of users' approval in those transactions. Those data marketplaces can be general-purposed ones such as Dawex or AWS Data Exchange [23] or can be specialized in precise industries. We can cite for example Veracity for energy and maritime transportation, Otonomo for vehicle data, or Quandl for financial and economic data.



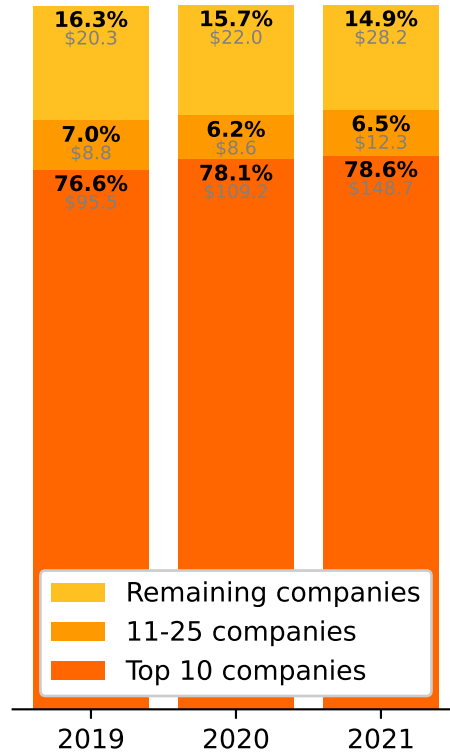
Source: IAB / PwC Internet Ad Revenue Report, FY 2021

Figure 1: Historical revenue (2011 – 2021) (\$ billions) [29]

A third of those data transactions are related to personal and marketing data. Digital advertising remains indeed the most lucrative source of revenue for internet companies with \$189.3 billion earned in 2021 (Figure 1), three quarters of which has been earned by 10 companies alone (Figure 2). Online searches remains the first advertising format and account for 41% of those revenues (Figure 3).

A tremendous amount of data is indeed generated by search engine users with increasing order of granularity. This data is then used profile user for improving targeted advertising (direct data monetization) and to personalized search results (indirect data monetization). Given the nature of search engines and their highly frequent use, this data, specially when cross-referenced with other sources of data such as social media applications, can become so comprehensive as to be considered as the online self-identity of a user. Some tech companies have been able to earn a massive amount of revenue through user profiling and thanks to decades worth of data processing expertise while users have been kept in the dark as to how and who accessed their data. Understanding how your personal data is processed is a task particularly difficult when the lifecycle of private data can go through complex and hidden layers of processing protocols and "invisible" data brokers transactions [30]. This creates a dismal power-asymmetry between what can be called data subjects and data controllers [11]. Recent legislative developments that have occurred in the western hemisphere, primarily the General Data Protection Regulation (GDPR) [32] in the EU and the California Consumer Privacy Act (CCPA) [27] in the US, attempted to mitigate those vulnerabilities and bring some balance to the data economy primarily regarding the privacy aspect. This brought major improvements but there is still a lot more to be done to truly empower data subjects in the current state of the digital economy.

No legislations have been put into place, for example, to enforce data controllers to compensate data subjects for their data or to set legal basis for data trusts [13]. Data trusts can be described as legal independent entities that would be bound by law and fiduciary obligation to exercise the rights of data subjects and negotiate the use of their data on their behalf. Personal information management and personal data monetization tools



Source: IAB / PwC Internet Ad Revenue Report, FY 2021
 Note: Totals may not equal 100% due to rounding

Figure 2: Three year (2019 – 2021) internet advertising revenue concentration (\$ billions) [29]

already exists [5,10,12,14,22,33] as well as centralized or decentralized data marketplace for personal data monetization [1,18], but they failed to focus on transparency and ergonomics and to propose a more extensive and permeable ecosystem. Existing solutions also remained focused on decentralized storing, which can raise some skepticism from current Big Tech companies regarding the scalability and safety of decentralized technologies. Most of those tools also only accept transactions made with specific cryptographic tokens and can only be operated on their in-house blockchain which can add another barrier of entry for companies and ordinary users.

We strongly believe that a better digital economy would follow the principles of data dignity and total data governance for the users as described by Jaron Lanier and E. Glen Weyl [25]. We, however, disagree with the authors on the inadequacy of using blockchains and decentralized technologies to create this new and empowering digital economy [24]. We think that both centralized and decentralized technologies could be used to implement an ecosystem that would support those principles and relieve the current vulnerabilities and asymmetries existing in the current state of this industry. To bring this vision to life, our ecosystem must rely on foundational in-house applications that would better support our mission. In this paper, we will consequently present the Khamseen Network and bring technical details for the first phase of our ecosystem which will be based on four different applications: a search engine letting users submit searches for Web2 and Web3 contents and giving them access to A.I generated insights, a hybrid data marketplace, a library of data trust modules giving users the freedom to choose with precision their privacy and monetization settings, and finally a more meritocratic and empowering ERC20 token designed to be both a store of value and an optional utility token.

2 Khamseen Network: a proposition for a more balanced data economy

The Khamseen Network is an organization born out of enthusiasm for the potential of Web3 and decentralized technologies and disappointment by its current applications. We believe that Web3 technologies have the maturity to support projects and applications that can become ubiquitous in the daily life of the regular internet user. However, it is also our belief that this technology exhibited some economic exuberance and became saturated by projects created for speculative endeavours and quick financials which could remind us the state of the web industry before the dot com bubble of the 1990's [16]. It is our conviction that decentralized technologies could instead be used to alleviate the detrimental trends rising from the centralization and opacity existing in Web2 technology and more specifically the alienating manipulation and monetization of private data in the current digital economy. This industry could indeed benefit from the transparency, safety and democratic nature of decentralized technologies. There are a few companies and organizations that already vowed to correct some of those harmful aspects through decentralized technologies and which partly inspired our current work. We can cite for example the Brave foundation which leveraged centralized and decentralized technologies for a better advertising economy and compensate users for their attention [31] or cheqd which implemented a decentralized ecosystem for self-sovereign online identity and digital trust [4]. Our mission, which is quite ambitious but also necessary, is to build a better digital economy that will assure that users, might they be individuals, research teams or companies remain master of their data and actually be financially and fairly compensated for it. We think that creating an ecosystem that will leverage the best of both Web3 and Web2 technologies to support that mission is necessary and would strongly alleviate the current asymmetrical nature of the relationship existing between data subjects and data controllers.

2.1 Search engine, private data and data trusts

Digital advertising represent a considerable source of revenue for tech companies and continues to increase every year, reaching \$189.3 billion earned in 2021 (Figure 1), a 35.4% increase from the year prior. Three quarters of those revenues have been earned by merely 10 companies (Figure 2) while online searches account for 41.4% of those same revenues i.e \$78.3 billion in 2021 (Figure 3). This already represents a colossal amount of money without accounting for the resell of user data generated from those online searches. Users receive virtually no financial compensation from those incomes and are actually in a vulnerable position against data controllers since their data is processed and sold in an increasingly complex and opaque digital infrastructure. Tech companies have indeed been investing a considerable amount of money and human expertise into data science and machine-learning / A.I for the sole purpose of user profiling and targeted advertising [15]. At the same time users remain overwhelmed by the extent to which their activity are being algorithmically recorded and scrutinized in spite of recent controversies [6] and legislative efforts such as the General Data Protection Regulation (GDPR) [32] in the EU and the California Consumer Privacy Act (CCPA) [27] in the US. Those new laws have given the ability to users to have easier access to their own collected personal information and to choose to have it deleted or not. More importantly, they now have a say in how their data is processed and can even object to the processing or the selling of their data in some cases. The application of those laws remains however lacking in user ergonomoy and it is our opinion that tech companies are not necessarily acting in good faith under that new legal framework.

We will leverage this new paradigm and go further into it to implement data trust modules and add monetization scheme for all users inside our ecosystem. Data trusts have been proposed by multiple experts [13, 25] to bring more power to data subjects by representing a set of users in any type of data transactions. They are defined as legal entities

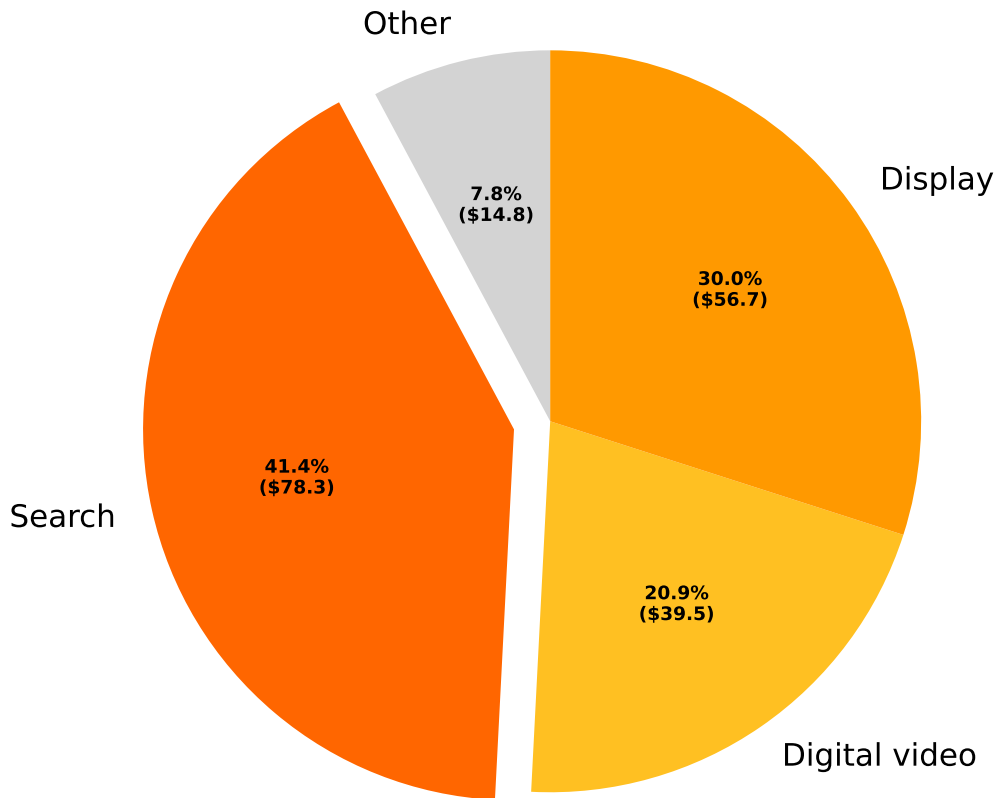


Figure 3: 2021 advertising format by share (\$ billions) [29]

that have a fiduciary and ethical duty to protect the users interest in the data economy according to an agreed upon set of rules and have the expertise to handle the legal and technical aspects of this industry. In order to really be efficient however, those data trusts must have a legal basis to enforce compliance from data controllers and truly protect data subjects. Unfortunately, no laws have been made to support the existence of data trusts. This is why we propose the implementation of "data trusts-like" modules inside our own ecosystem that will be enforced by blockchain smart contracts.

Inside the Khamseen Network ecosystem, users will have complete access to their personal data that have been collected and will be able to delete it or extract it at any time automatically from the platform. They will be free to choose what information they allow to be collected or monetized. They will also be able to blacklist or whitelist companies, individuals and organizations registered as data buyers in our ecosystem. Given the versatile nature of the data collected and the multiple ways it can be processed or leveraged, it can be unclear for users what part of the data they are comfortable sharing and monetizing and under what terms [13]. Making a choice can therefore become a daunting task for a user. This is why a set of prebuilt data trust modules will be created to help users navigate the fog from their side of the data economy. They will have of course the ability to opt-out of a data trust module at any time and choose to register to another one, create their own data trust module or customize for themselves their own privacy and monetization settings. Inside our own ecosystem, different data trust modules will indeed coexist and compete with each other to help users choose the type of data policy they are willing to accept. Since users will get royalties from the sell of their data and since the quantity and quality of this data will dictate the amount of royalties they will get, data trust modules will help the community find their own balance between privacy and monetization.

Applications will continuously feed data to those modules throughout the different

phases of the Khamseen Network ecosystem. The first application will be a search engine, since it accounts for a major part of the revenues from digital advertising. This search engine will make use of users data for user profiling to support the recommendation system and the advertising infrastructure based on sponsored links and personalized ads. This will create revenues that will be put back into the different data trust funds to be finally distributed among their beneficiaries. This search engine will also incentivize users to generate themselves high quality data such as product reviews for example through additional reward schemes. Users will therefore have multiple streams of data income besides their own private data monetization (Cf. 3.1.5)

2.2 Data marketplace

We believe that our applications shouldn't be the only source of data for potential data buyers inside our ecosystem and that access to a broad range of data from different type of industries and fields of study will be useful in our platform. Developing a data marketplace inside our ecosystem will allow data buyers and vendors to transact in a more transparent, fair and secure environment. This data marketplace will support our mission of creating a better digital economy by avoiding the obscure data brokers and hidden and secretive data transactions that rid the current industry and help us better negotiate and broadcast our own data offers. This data marketplace will be built with a similar layered approach than most of modern data trading entities [3] (Figure 4) and will also be based on a hybrid architecture incorporating centralized and decentralized mechanisms. This will allow both fiat and cryptocurrency payments (Cf. technical details in 3.1.4) to be made inside the marketplace as well as giving the ability to data vendors to store proprietary datasets in our own centralized servers or to make them available on decentralized servers (IPFS). This data marketplace will also provide tools to allow any third-parties to easily and quickly create a paid or free-to-use customized data access API for their data and allow buyers to connect.

Data processing tasks could also be requested on this data marketplace by any organizations in order to curate, label or clean data points. We will let requesting organizations decide beforehand if and how they wish to compensate contributors. If participants are indeed compensated, the profits will be distributed among them according to their own contributions (more details will be shared at the appropriate time in accordance with the requesting organization requirements and guidelines). Data processing tasks can also be requested by Khamseen Network developers and in that case a compensation scheme will always be put into place to reward users who are willing to participate. The rewards mechanism will depend on the difficulty or complexity of the task and will explicitly be defined beforehand for each project. Users that have put effort into the data processing project will continue, in some cases, to earn royalties. Those data processing projects will of course be optional and will not in any case disadvantage users that choose to not participate.

2.3 Why another token?

The Khamseen token is first and foremost a showcase of our principles of openness and user empowerment applied to the conception of an ERC20 token. We designed this token to be a store of value first, secondly to be an optional utility token for our ecosystem and lastly to bring awareness to our mission. Since this token will have a relatively low supply for an utility token, we don't plan to exclusively aliment our ecosystem by this token. We remain open to the possibility of accepting other ERC20 tokens, cryptocurrencies or fiat money for our paid services. Since this token is also an additional way of supporting the effort of this organization, Khamseen token holders will be rewarded with the highest priority for the different purchases and transactions in the ecosystem. This includes purchasing advertising space or sponsored links on the search engine, commissioning data processing tasks and paying for data access.

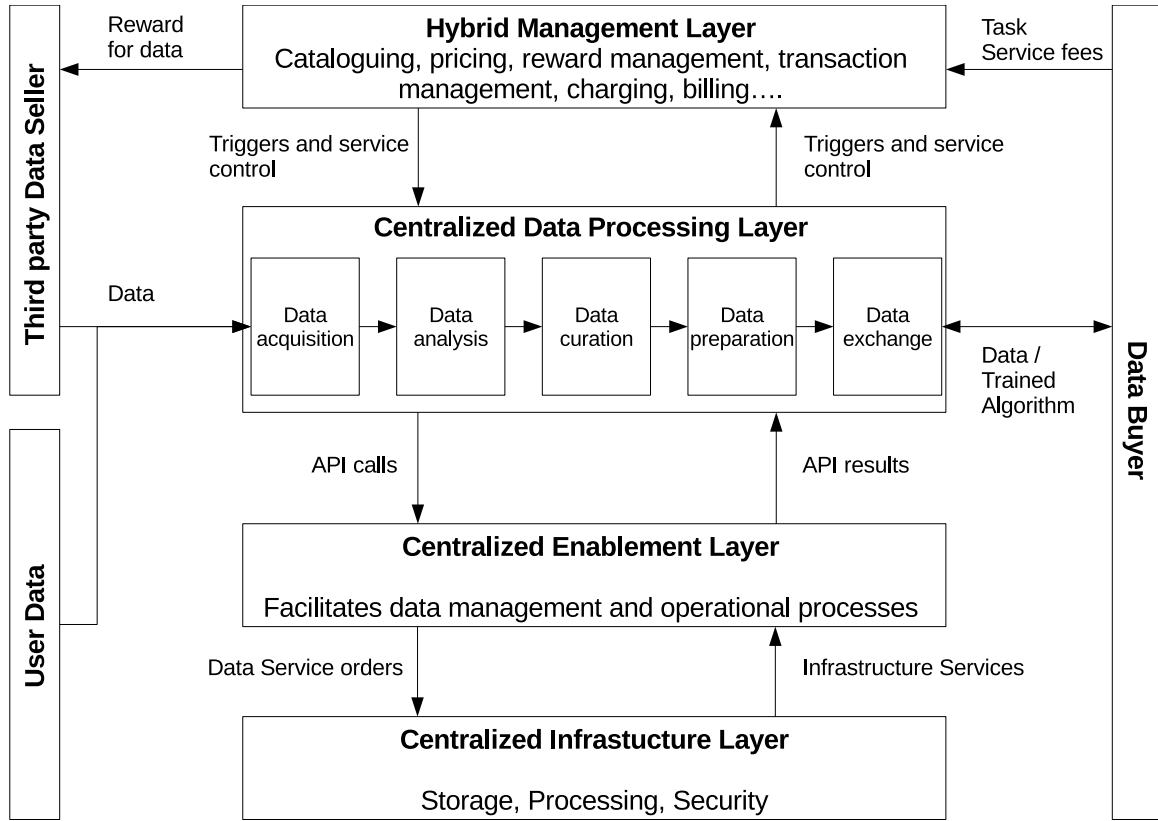


Figure 4: A layered and hybrid approach to data trading

3 Technical aspect of Khamseen Network Ecosystem v1.0

3.1 Khamseen Network Data Platform

The first phase of the Khamseen Network ecosystem will be based on four different applications linked with each other: a library of data trust modules, a search engine for Web2 and Web3 contents, a hybrid data marketplace and finally a more meritocratic and empowering ERC20 token. The order of deployment is described in our roadmap (Cf. 4.). Those four applications designed to support each other in a virtuous circle will allow us to build a harmonious ecosystem to better leverage data, empower users and make a step towards a more balanced and transparent data economy.

3.1.1 Data trust modules

The Khamseen Network platform will offer a series of pre-built data trust modules and implement a tool available to the community to create their own modules. Users will have the possibility to register to one or multiple data trust. Since a data trust module will correspond to a set of specific datasets pooled from the beneficiaries according to a specific monetization policy, the profits generated from the sell of the datasets will only benefit members of the trust. The Khamseen Network developers will take care of indexing and listing the different trusts and datasets in order to help users make a choice as well as to prevent redundancy and automate processes related to the monetization of those datasets. Data trust modules can be established in the Khamseen Network ecosystem in a partially centralized or in a fully decentralized fashion. In the case of the partially centralized method, a simple database will record the beneficiaries and link the affiliated datasets. This is the simplest and cheapest method to join a data trust module. If some users want to fully leverage the blockchain technology to better protect and enforce the policies of their data trust, a smart contract template will be available so that any user can deploy on their own

a data trust smart contract on the Ethereum blockchain. This smart contract will follow the ERC-721 standard and will contain in its metadata at least the following information:

- the name of the data trust
- the Khamseen Network API endpoint returning a file describing the policies of the data trusts (in txt or json format)
- the Khamseen Network API endpoint returning the list of datasets generated by the trust
- the addresses of each smart contracts linked to the datasets affiliated to the trust (more details in 3.1.4)

User will be able to register or unregister to the data trusts through a simple public function inside this smart contract.

3.1.2 Hybrid search engine

The Khamseen Network search engine will be designed as a hybrid search engine alimeted by multiple blockchain scanners monitoring the most popular blockchains and web crawlers deployed on both Web2 centralized servers as well as Web3 decentralized servers, mainly IPFS nodes.

Running a performant and reliable search engine can be very costly in bandwidth and computational power. In order to accelerate development and reduce costs for the first phase of our network, we plan on building an alpha version of the search engine restricted to a set of core themes: cryptocurrency, finance, technology, art, science, news and entertainment. We plan however to add more themes according to the needs and demands of users as well as resources available to the Khamseen Network developers in the next phases of the ecosystem. The Web2 crawlers that will alimted the search engine will consequently browse and index pages from domains that fall under those specific themes. At the same time, the IPFS web crawlers will be connected to a backend referencing engine in order to automatically classify, filter and index the appropriate content into easily accessible and human-readable addresses on our own servers. The Khamseen Network will be an active participant of the IPFS protocol and run multiple nodes to support the needs of our own ecosystem. Finally, blockchain scanners will offer the possibility for users to easily access statistics from the blockchains and recent transactions.

This search engine will indeed differ from other popular search engines and will provide processed and aggregated results on top of indexed pages and content (Figure 5). Different AI-based models will process, aggregate, and classify contents to offer relevant insights and intelligence regarding the search results. Those insights will be of different nature, depending on the subject of interest for the user. Sentiment classification can be an example of natural language processing task that will be incorporated into the search engine results, allowing users to extract the broad sentiment on a specific subject or to rank search results according to the positivity or negativity of their content. Quantative analysis on cryptocurrencies is also an example of the type of insights that could be brought by the search engine. By aggregating statistics extracted from blockchains with articles and data from the web it will be possible, for example, to quantify the overall interest on a specific cryptographic token or a DeFi project.

By leveraging centralized and decentralized content and data in a synergistic way we will allow our users to have a better view of Web2 and Web3. This will also help our developers in their mission of user empowerment by providing better insights and "web intelligence" and creating new services for our userbase.

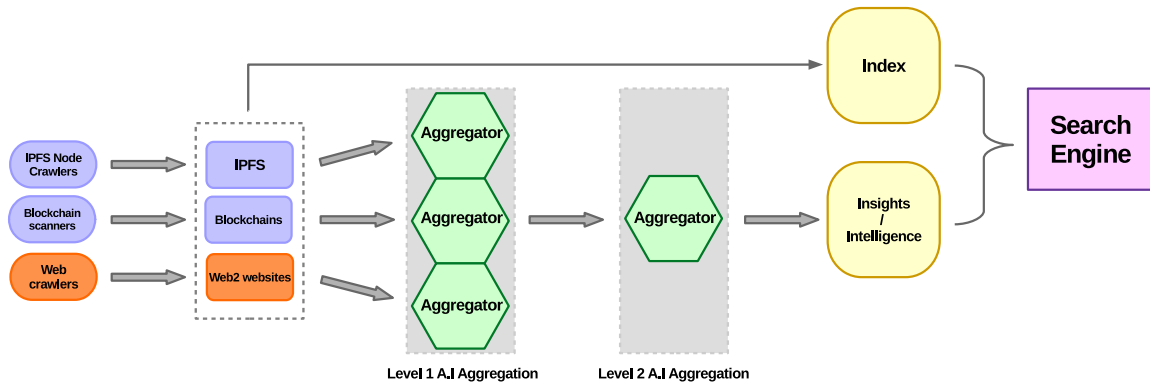


Figure 5: Khamseen Network’s hybrid search engine

3.1.3 Digital advertising

For the first phase of the ecosystem, two marketing tools will be available to the advertisers: frontpage ads and keyword-based sponsored links on the search engine. While frontpage ads will be purchasable on the Ethereum blockchain, keywords sponsored links purchases will only be accessible in a centralized way through the digital advertising portal of the Khamseen Network platform for at least the first phase of the ecosystem.

A dedicated oracle smart contract deployed on the Ethereum blockchain will make different ad placements available for purchase to the advertisers (Figure 6). An automatic first price sealed bid auction mechanism implemented on this smart contract will allow advertisers to bid for the ad placement of their choice. The advertisers will simply need to call the bidding function, precise the following arguments:

- the ID of the ad placement as documented in our marketing tool documentation
- the number of days they are willing to rent the ad placement for, between one and thirty days
- the links for the advertisement asset files (which could be hosted by the Khamseen Network if needed)

and then send the amount of Khamseen tokens or Ether the advertiser is willing to pay. Since we offer the possibility to bid with Khamseen tokens and/or with Ether, the Khamseen-Ethereum mean exchange rate of DEXs will be applied.

During a first price sealed bid auction, participants can only bid once and cannot know what are the other bids. For each ad placement available, advertisers can bid during the 24 hours starting from the first bid and precise the number of days they wish to rent the ad placement. At the end of the auction, the bidder with the highest price per day wins the auction. If there is a tie, the bidder bidding with the most amount of Khamseen tokens wins the auction. The unsuccessful bids are then refunded to their respective bidders. At the end of the rented time, the auction is opened once again but the previous buyer’s ad will continue to be displayed on the search engine until a new auction winner comes up. The search engine will scan this smart contract on a daily basis and, for each ad placement, eventually close the auction, select the winners and retrieve their advertising assets. The Khamseen Network developers have a right to veto ads if they infringe on the community guidelines which will be defined at the appropriate time on the Khamseen Network platform.

3.1.4 Data marketplace

The Khamseen data marketplace will give to any user of the platform, whether they are an academic research department, a private company, a non-governmental organization or a

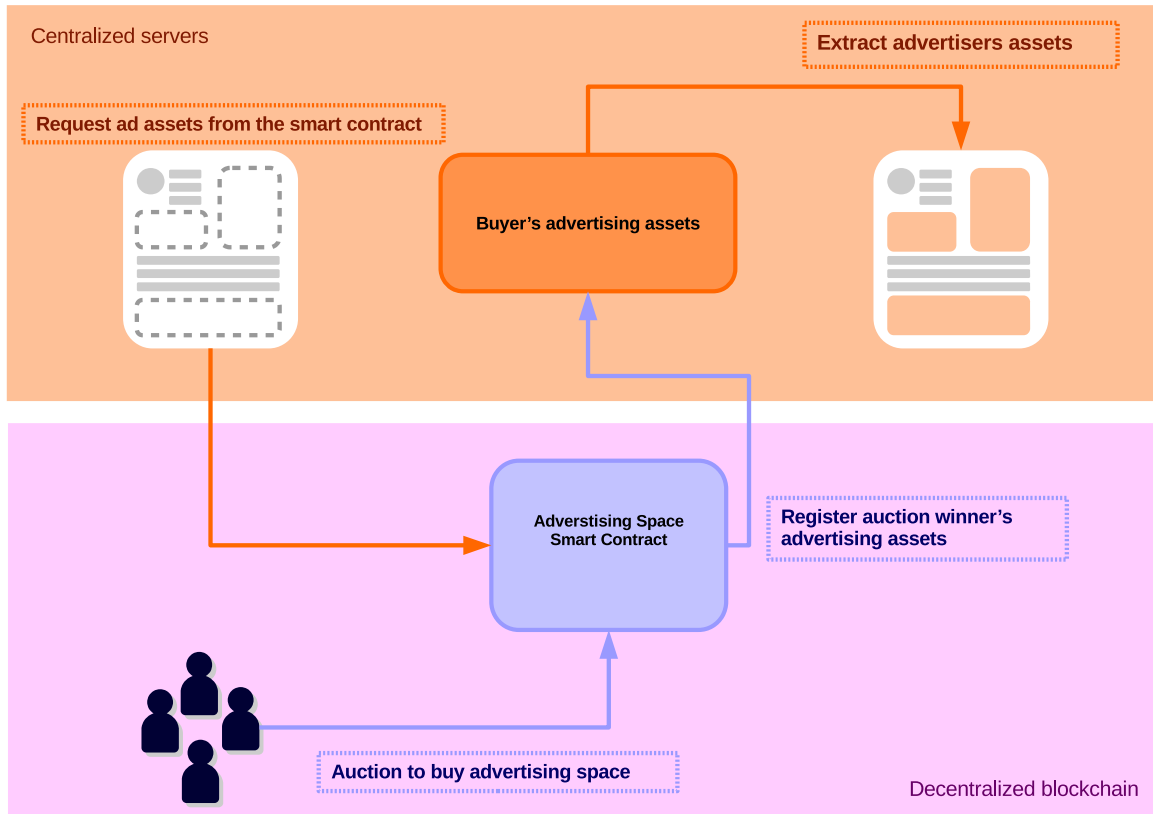


Figure 6: Decentralized purchase for digital advertisement

single individual, the ability to protect their proprietary data behind a paywall or to make it freely accessible and to define their own customized data access API and pricing tiers.

The data marketplace will be built on an architecture incorporating both centralized and decentralized mechanisms for data storing and payments. For the first phase of the platform, proprietary data with paid access will be stored in our encrypted and centralized servers and will only be accessible through a specific API while freely accessible data will be either stored those same servers or on the IPFS protocol. Data transactions will also be made possible on the Ethereum blockchain, with Ether or any tokens data vendors are willing to accept, as well as on our centralized platform in order to process fiat money payments.

The decentralized part of the data marketplace infrastructure will make use of oracle-based smart contracts for handling communication between the blockchain and our centralized platform API. These oracle smart contracts are what we can refer to as database smart contracts which will mainly register purchases, grant or restrict access to data consumers and handle the compensation logics.

Data vendors, whether they are a third-party, a set of users in a data trust, or the Khamseen Network itself, will mainly use the centralized part of the architecture, and more particularly the data vendor tools that will be accessible on the platform website. Those tools will allow vendors to easily create a vendor account, upload the data, add any additional information related to their data, define the data access API with its quotas and restrictions, and submit their pricing offers and the list of accepted currencies. If only fiat money are accepted by the vendor, the transactions will be simply handled in our centralized servers. However, if they are willing to accept purchases made in cryptocurrencies, or they simply wish to bring more transparency to those purchases, they will have to add the wallet address to wish to receive the funds from, and cover the cost of deployment of a database smart contract during registration. This smart contract will then be automatically built according to the data vendor commercial offer and deployed on the blockchain by

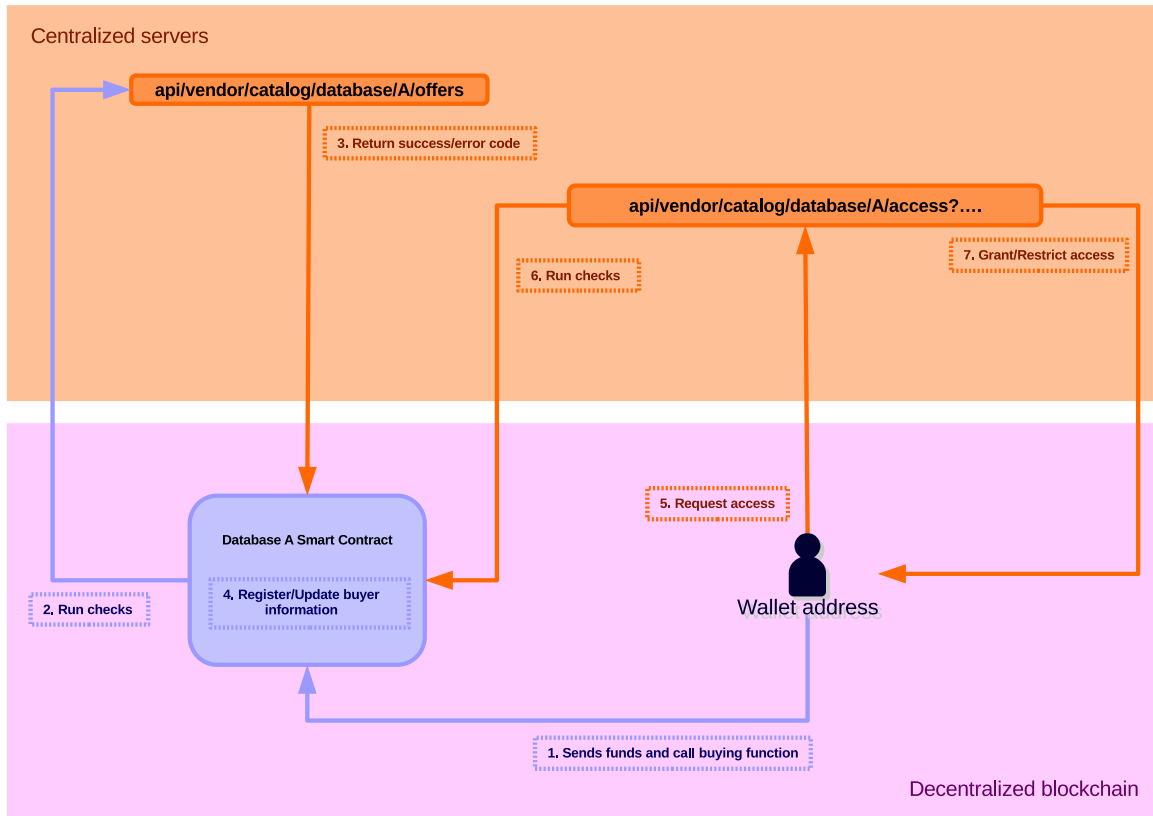


Figure 7: Decentralized data purchase

our platform. The figure 7 describes how a decentralized data purchase is made. The database smart contract will contain a function that will allow buyers to get access to the data and eventually choose which offer made available by the vendor they wish to purchase. This function will then call the platform API endpoint corresponding to this specific database commercial summary in order to check that the buyer has correctly asked access to an existing pricing option for the database and already sent the appropriate amount of cryptocurrencies. If all checks have been successful, the purchase will be added to the vendor historic in our centralized server while the database smart contract will register the purchase by creating or updating a mapping variable containing the buyer address, the date of the purchase and the corresponding pricing tier. This is that variable that will be scanned by our centralized backend engine to grant or restrict access to the data access API for buyers that transacted on the blockchain. The database smart contract will also contain a function that will handle the distribution of the profits to the vendor (Figure 8) and possibly the transfer of service fees to the Khamseen Network for each purchase). If the database has been built through the contribution of users, this same function will allow user to get their rewards and/or royalties (Cf. 3.1.6).

Data vendors will have the possibility to set their own price according to their own business model and have the choice, for example, to set a fixed price, a usage-based-price depending on the number of calls or a volume-based price depending on the amount of data requested by the buyer. They will also have the freedom to set up a subscription-based or a one-time payment model. Since it remains a difficult task to price an asset such as data [28] due to its non-depletable and non-excludable nature, the Khamseen Network experts will be available to help vendors set up a pricing model if needed. Data vendors will also be encouraged to make available free samples of their data in order to help buyers reduce their risk of uncertainty regarding the relevance of the data to their current needs. This is specially crucial for buyers interested in training an AI/ML model and comparing which dataset would potentially lead to better predictive performance [2].

3.1.5 Data processing task requests

The Khamseen Network data marketplace will also be the place where manual data processing projects can be requested by any third-party or by the Khamseen Network itself in order to be sold or to support our own R&D department and potentially create additional services and commercial applications into our ecosystem. This manual data processing portal available on our platform will let users manually label data or help in other data processing tasks and earn a reward and possibly royalties according to the quality and quantity of their contribution. The precise compensation scheme will always be explicitly described before each projects for better transparency.

The rewards are distributed at the end of each project, when all the data points have been successfully processed by enough users and when the project predefined milestones have been completed. For each data points, a consensus result is constructed and will be regarded as the "correct result". This consensus result can be computed in very different ways depending on the type of data and is only constructed in order to quantify each user contributions. It can differ from the method chosen in production by the project commissioner or by the Khamseen Network developers for downstream tasks. A "task score" is then computed for each contributor according to the formula described in (1).

$$Task\ score = \max(1, \text{Number of "correct results"} - \text{Number of "errors"}) \quad (1)$$

This "task score" is then used to compute the actual reward and eventually royalties earned by each contributor according to the formula in (2).

$$Reward = Task\ Score \times \frac{Sell\ Price \times User\ Share}{\sum Contributors\ Task\ Scores} \quad (2)$$

3.1.6 Contributors and compensation scheme

The compensation logics for data contributors, whether they are members of a data trust or a participant of a data processing project will be mainly handled by each database smart contracts (Figure 8). If a user wishes to retrieve his earnings from a single database smart contract, he will simply have to call a specific function implemented in this smart contract. This function will call the user and database API endpoints on our platform to extract the reward/royalty formula, the number of purchase, the contribution and compensation metrics of the users in order to compute the latest amount earned. This function will then directly transfer the amount to the user, register this transfer and update the user compensation metrics on the backend engine and smart contract to prevent double spending. Most of the computing and logics will be kept off-chain during the first phase of the platform in order to reduce the costs of transaction on the blockchain, but we do not exclude migrating more processes and storing metrics on-chain in the future if a more decentralized architecture would be preferred by the community. All the metrics, formulas and information used to compute the earnings will be openly and easily accessible on our centralized platform to make the compensation mechanism as transparent as possible. The amount to be earned will therefore be known at all time for each user so that they can decide whether that amount is profitable against the current blockchain transaction fees.

A compensation smart contract will also be created and deployed to wrap the compensation logics from multiple database smart contracts to allow users to extract their rewards, royalties and other earnings made from data sold on the marketplace without the need to transact with each database smart contract (Figure 9). This compensation smart contract will contain a function that will call the user API endpoint in order to first extract the list of all datasets and databases the user has contributed to and their respective database smart contract addresses. Then, for each of those databases, the procedure remains the same: the user and database API endpoints on our platform will be called to extract the needed

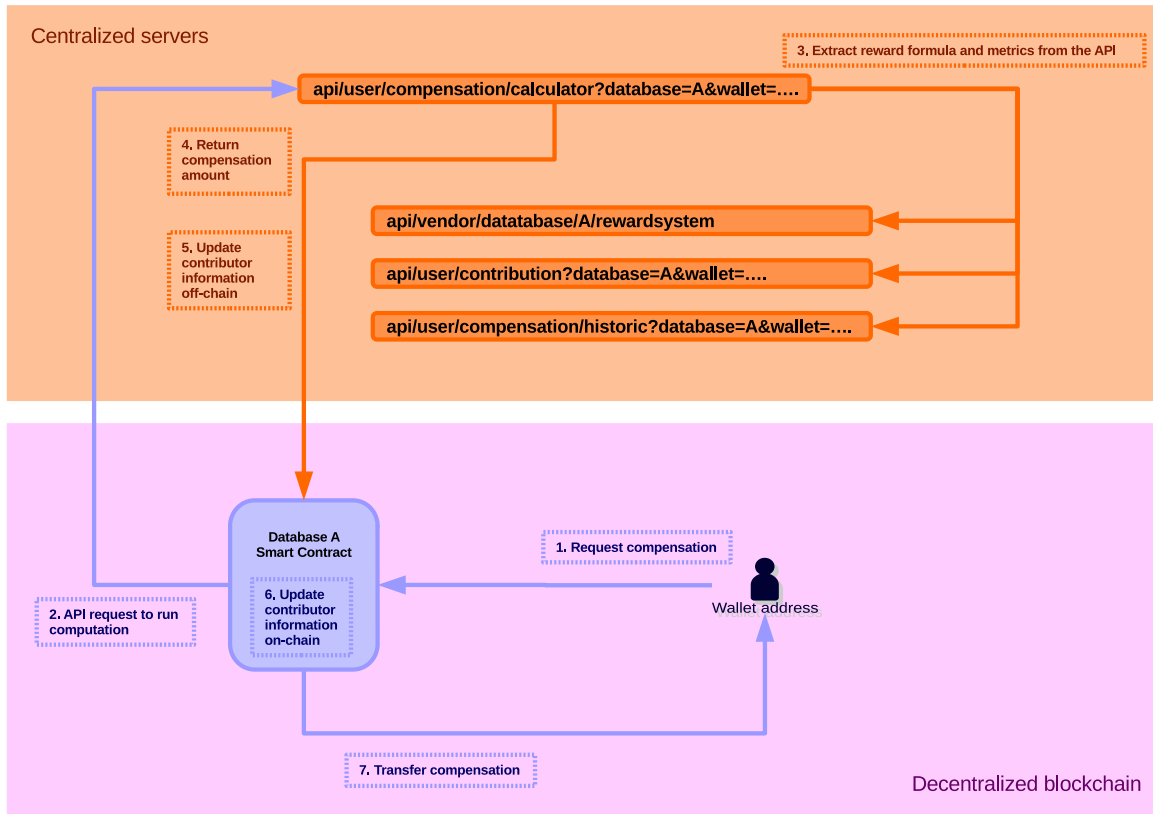


Figure 8: Decentralized compensation mechanism

data in order to compute the latest amount earned. The compensation smart contract will then call the different database smart contracts to approve the transfer of cryptocurrencies from the compensation smart contract to the user wallet address. It is important to note that the compensation smart contracts will not store the funds, whether it is Ether or any cryptographic token, but will be automatically approved allowances for each purchase from the related database smart contracts which will actually store the funds. This is why for the first phase of the network, only ERC20 tokens that have a functional approval mechanism will be accepted for purchasing data that has been built through user contribution and that have a defined compensation scheme.

When datasets and databases built through user contribution have, on the other hand, been purchased with fiat money, they will automatically be converted to Ether or Khamseen tokens and send to the database smart contract in order to compensate contributors. We exclude indeed, at least for the first phase of our platform, directly compensating contributors in fiat money.

3.2 Khamseen Token

3.2.1 Token summary

The Khamseen token is a mintable Ethereum ERC20 token with a cap set to 50 million and a number of 16 decimals. A tenth of that cap i.e 5 million tokens have been preminted at the time of the deployment and immediately transferred to a capped crowdsale smart contract. No preminted tokens have been attributed to any other wallet and the crowdsale smart contract has been designed to not be able to transfer back tokens to any other wallet as to prevent any fraudulent schemes, which is crucial when a compound effect has been implemented in the token minting process (Cf. 3.2.2). The cap value of 50 million has been arbitrarily chosen in order to have a reasonable number of tokens circulating and avoid creating a token that is so rare that it could end in the hand of a fortunate few or so

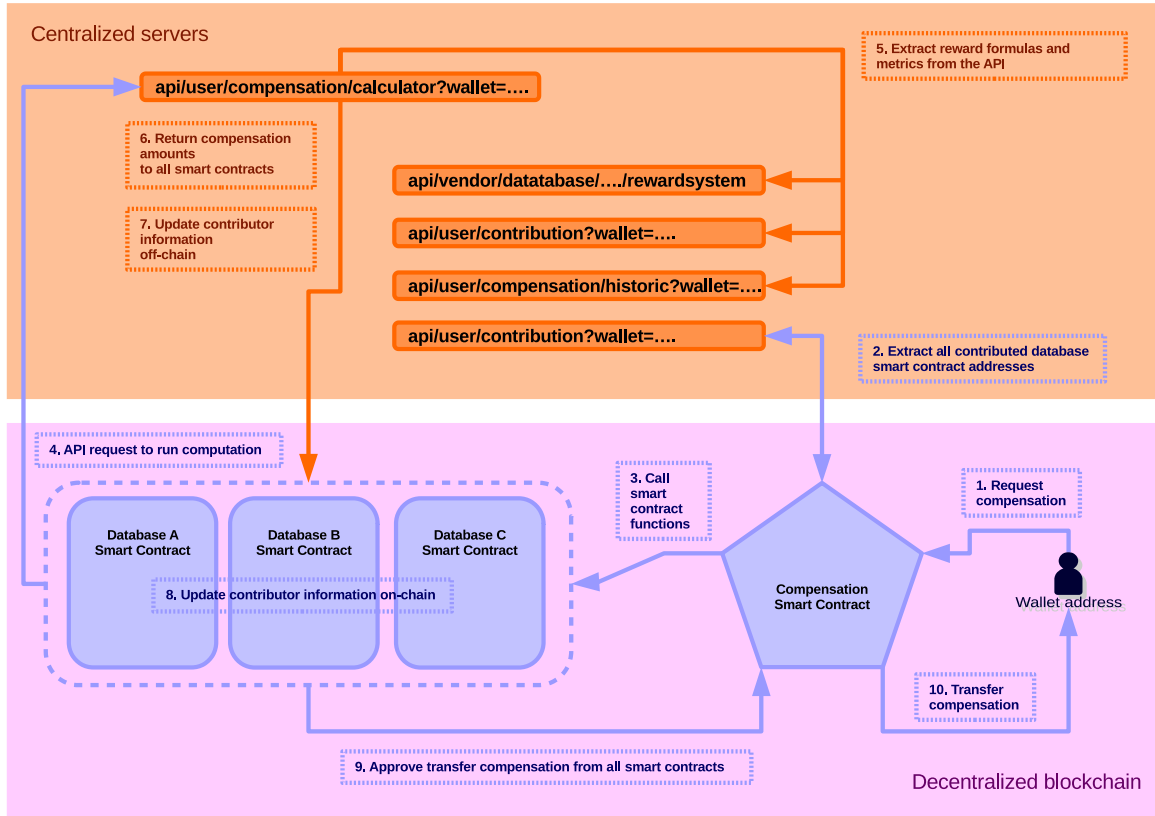


Figure 9: Decentralized compensation mechanism

abundant that it would not have any substantial value. This token has indeed been designed to be foremost a store of value and secondly an optional utility token for any transactions that can be made inside the Khamseen Network ecosystem. The number of 16 decimals has been chosen to allow one token bit to be exchangeable against one single wei while setting the definitive crowdsale price of a token to 0.01 Ether. This allows wallets with small Ether balance to still be able to buy and earn newly minted tokens and participate in our ecosystem without needing to spend 0.01 Ether in one single transaction.

3.2.2 Minting mechanism and compound effect

There are only two ways to earn newly minted Khamseen tokens. The first method consists in minting directly by calling the "mintMinerReward" function on the smart contract while the second method consists in being the beneficiary address of a "mintMinerRewardForAddress" function call. Each time one of these functions is called, the message sender for the "mintMinerReward" function or the beneficiary address for the "mintMinerRewardForAddress" function get their minting streak incremented and are rewarded a specific amount of token. This amount is computed by a reward function inside the token smart contract which is described by the formula in (3).

$$Minting\ Reward = 10^{-\lfloor \frac{Total\ Supply}{10E6} \rfloor + 2} \times \max(1, \lfloor \frac{balance + streak}{2} \rfloor) \quad (3)$$

This reward function formula (3) simply consists of a guaranteed reward and a multiplier. The multiplier is the arithmetic average of the user balance and their minting streak. This multiplier has been designed as such to temper the exponential power of the compounding effect, to avoid reaching the maximum cap too quickly as well as to reduce the advantage of wallets with high purchasing power. This formula creates a continuous compound effect for the earnings [8]. The more token a user already own and/or the more frequent they mint the more they can earn tokens. The compound effect is a virtuous circle that exists for some fiat

money financial instruments but remain hardly accessible and has not been implemented in a wide fashion in cryptocurrencies and DeFi projects besides staking mechanisms which can involve substantial risks [9]. The possibility to earn new Khamseen tokens is therefore not related to the computing power available to the user nor their willingness to put liquidity at stake. With the guaranteed earnings we allow even the most modest wallets to participate in such mechanism with no risks and no cost of entry besides the Ethereum blockchain transaction fees. Another notable difference is that, here, the compound effect has no losers. There is no interest to be paid by any party. This reward mechanism has been designed to be easily computed, it has a low and constant complexity and it is not based on a constant competition between users. Therefore, it does not directly disadvantage newcomers. An empty wallet would indeed get the same guaranteed reward according to the current supply independently of the number of concurrent miners and their own liquidity.

A simple difficulty mechanism has been implemented in order to also counteract the exponential nature of the compound effect and delay the time needed to reach the maximum cap as to allow more newcomers to participate and become token holders. The difficulty mechanism is simple and consists in dividing the guaranteed reward term by 10 every 10 million token minted. The difficulty consequently increases in 5 steps instead of a continuous manner and initializes the guaranteed reward to 0.01 Khamseen tokens. The reason behind this choice, besides the simplicity of the computation is to allow users to predict with exactitude their rewards which can be crucial for proper financial planning, a characteristic needed in the current volatile state of DeFi projects. A function has been added into the smart contract to compute precisely the expected reward and make it easier for users to decide whether to call the reward functions and spend the blockchain transaction fees. The second reason behind this discrete, and not continuous, difficulty mechanism is to avoid reaching a state where rewards would become so low that it would not justify the cost of transaction fees and that the token would never reach its cap.

3.2.3 Token crowdsale

The proceeds earned from the crowdsale will help support development and the operational costs of the Khamseen Network but the success of the crowdsale is not a condition for roadmap completion and simply remains a mean to broadcast our mission and therefore, possibly accelerate and ease the development process. The token is first to be used as a store of value that can be earned without any privilege or consequential investment beside standard Ethereum blockchain transaction fees. One token can be obtained through this smart contract for a fixed and definitive price of 0.01 Ether. The crowdsale is open for all, is not time locked but only capped to the amount of wei corresponding to the amount of preminted tokens: 5×10^{22} wei. No limit has been set as to how many tokens a wallet can purchase in order to allow potential pooling communities to organize and operate inside our ecosystem.

4 Roadmap

The current goal of the Khamseen Network is to build an ecosystem of innovative applications and tools that will bring more balance and user empowerment to the digital economy. We presented here, only the first phase of the Khamseen Network ecosystem. We plan to bring life to our current vision for this ecosystem, to continue improving it and to integrate many more applications in the next phase. We plan, for example, on extending data monetization and applying data trusts modules for personal data generated outside our ecosystem as well as implementing a governance system for the community to submit proposals and vote on how to update and improve any processes inside our ecosystem. The costs and transaction fees incurred to the userbase inside our ecosystem will also be continuously monitored. We do not exclude the possibility of deploying our own blockchain and

side chains compatible to other popular blockchains in order to lower costs for users to a minimum and pushing towards more decentralization while still guaranteeing the integrity and transparency of processes taking place inside the ecosystem.

Roadmap:

- Q4 2022: deployment of the Khamseen token and start of the crowdsale
- Q2 2023: deployment of the alpha version of the Khamseen data marketplace
- Q3 2023: deployment of the alpha version of the search engine
- Q4 2023: deployment of the data trust modules
- Q2 2024: deployment of the beta version of the Khamseen data marketplace
- Q3 2024: deployment of the beta version of search engine
- Q4 2024: deployment of a governance system to help define the second phase of the Khamseen Network ecosystem

5 References

References

- [1] AIRBLOC. The privacy-first data exchange platform, 2018.
- [2] Santiago Andrés Azcoitia and Nikolaos Laoutaris. Try before you buy: A practical data purchasing algorithm for real-world data marketplaces. *arXiv preprint arXiv:2012.08874*, 2020.
- [3] Santiago Andrés Azcoitia and Nikolaos Laoutaris. A survey of data marketplaces and their business models. *arXiv preprint arXiv:2201.04561*, 2022.
- [4] Ankur Banerjee. The business models of identity, 2021.
- [5] BIGtoken. How to earn money with bigtoken, 2020.
- [6] Carissa Boerboom. Cambridge analytica: The scandal on data privacy. *Augustana Center for the Study of Ethics Essay Contest*, 2020.
- [7] Gabriella Cattaneo, Giorgio Micheletti, Mike Glennon, Carla La Croce, and Chrysoula Mitta. The european data market monitoring tool, 2020.
- [8] JAMES CHEN. Continuous compounding definition and formula, 2020.
- [9] Lin William Cong, Zhiheng He, and Ke Tang. Staking, token pricing, and crypto carry. *Available at SSRN*, 2022.
- [10] DATACOU. The personal data revolution, 2020.
- [11] Paul De Hert and Serge Gutwirth. Privacy, data protection and law enforcement. opacity of the individual and transparency of power. *Privacy and the criminal law*, pages 61–104, 2006.
- [12] Isabella de Michelis di Slonghella and Luca Bolognini. An introduction to the right to monetize (rtm), 2018.
- [13] Sylvie Delacroix and Neil D Lawrence. Bottom-up data trusts: disturbing the ‘one size fits all’ approach to data governance. *International data privacy law*, 9(4):236–252, 2019.

- [14] digi.me. A new ethical and sustainable way for individuals to take control of their data and privately share with data-driven apps and service, 2021.
- [15] Christopher Ifeanyi Eke, Azah Anir Norman, Liyana Shuib, and Henry Friday Nweke. A survey of user profiling: State-of-the-art, challenges, and solutions. *IEEE Access*, 7:144907–144924, 2019.
- [16] Yuchao Fan. Dissecting the dot-com bubble in the 1990s nasdaq. *arXiv preprint arXiv:2206.14130*, 2022.
- [17] Josh Gottlieb and Khaled Rifai. Fueling growth through data monetization, 2017.
- [18] Roger Haenni. White paper v15. *Source: <https://datum.org/assets/Datum-WhitePaper.pdf>*, 2019.
- [19] Elizabeth Heichler and Munther Dahleh. Why the data marketplaces of the future will sell insights, not data. *MIT Sloan Management Review*, 60(1):1–4, 2018.
- [20] Nicolaus Henke, Jacques Bughin, Michael Chui, James Manyika, Tamim Saleh, Bill Wiseman, and Guru Sethupathy. The age of analytics: Competing in a data-driven world, 2016.
- [21] Dr. Moritz Holm-Hadulla, Dr. Christian Hamann, Dr. Hannah Bug, and Nora Johanna Meiborg. Data 360° – data pooling between companies, 2022.
- [22] HU-MANITY. We believe your data is your property, 2021.
- [23] Douglas B. Laney. Data monetization: New value streams you need right now, 2020.
- [24] Jaron Lanier. *Who owns the future?* Simon and Schuster, 2014.
- [25] Jaron Lanier and E Glen Weyl. A blueprint for a better digital society. *Harvard Business Review*, 26, 2018.
- [26] Mondato. Data is the new dogma: Monetize and thrive, 2020.
- [27] State of California. California consumer privacy act of 2018, 2018.
- [28] Jian Pei. Data pricing—from economics to data science. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 3553–3554, 2020.
- [29] US Pwc. Iab internet advertising revenue report-full year 2021. Technical report, Technical Report April, 2022.
- [30] Leanne Roderick. Discipline and power in the digital age: The case of the us consumer data broker industry. *Critical Sociology*, 40(5):729–746, 2014.
- [31] Brave Software. Basic attention token (bat) blockchain based digital advertising, 2021.
- [32] European Union. General data protection regulation (gdpr). regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation) (text with eea relevance), 2016.
- [33] VETRI. Vetri whitepaper, 2017.

6 Important Information

This Whitepaper is provided by Khamseen Network for informational purposes only. Nothing in this Whitepaper shall be construed as an offer to sell or buy securities in any jurisdiction, or a solicitation for investment, or an investment advice. The Whitepaper does not regulate any sale and purchase of Khamseen tokens (as referred to in the Whitepaper). The use of Khamseen Network applications is subject to the Platform Terms and Policies.

This Whitepaper describes the current vision for the Khamseen Network. While we intend to attempt to realize this vision, please recognize that it is dependent on a number of factors and subject to risks. It is entirely possible that the Khamseen Network platform will never be implemented or adopted, or that only a portion of our vision will be realized. We do not guarantee or warrant any of the statements in this Whitepaper, because they are based on our current beliefs, expectations and assumptions, about which there can be no assurance due to various anticipated and unanticipated events that may occur. Some aspect of the technologies used for the Khamseen Network platform is in its infancy and will be subject to many challenges, competition and a changing environment. We will try to update our community as things grow and change, but undertake no obligation to do so.

Due to the retrospective nature of regulatory action or guidance, we can make no guarantees re-garding the legality of Khamseen Network or the Khamseen token launch in any given jurisdiction. We must operate Khamseen Network in accordance with the laws of relevant jurisdictions. As such, Khamseen Network or Khamseen tokens may not be immediately available in certain countries. Khamseen tokens are non-refundable and are not for speculative investment. No promises of future performance or value are or will be made with respect to Khamseen tokens, including no promise of inherent value, no promise of continuing payments, and no guarantee that Khamseen tokens will hold any particular value. Khamseen tokens are not securities and are not a participation in the Khamseen Network company or foundation (once established). Khamseen tokens hold no rights in the company or foundation (once established).

This Whitepaper may be updated or altered, with the latest version of the Whitepaper prevailing over previous versions and we are not obliged to give you any notice of the fact or content of any changes. The latest version of the Whitepaper in English is available at www.khamseennetwork.org/assets/pdf/khamseen_white_paper.pdf. While we make every effort to ensure that all data submitted in the Whitepaper is accurate and up to date at the point in time that the relevant version has been disseminated, the proposed Whitepaper is no alternative to consulting an independent 3rd party opinion. The Whitepaper does not constitute an agreement that binds the Khamseen Network company or foundation (once established). The Khamseen Network company's or foundation's (once established) directors, officers, employees and associates do not warrant or assume any legal liability arising out of or related to the accuracy, reliability, or completeness of any material contained in the Whitepaper. To the fullest extent permitted by any applicable law in a ny jurisdiction, the Khamseen Network company or foundation (once established) disclaim all liability to you and everyone else in respect of the content of this Whitepaper, whether under any theory of tort, contract or otherwise and whether in respect of direct, indirect, consequential, special, punitive or similar damages. Persons who intend to purchase Khamseen tokens, should seek the advice of independent experts before committing to any action, set out in the Whitepaper.