Koi

Robust Incentives for Content Curation in the New Web

Alexander Morris
al@openkoi.com

Sam Williams
sam@arweave.org

Abstract

The Koi Network provides incentives for participants to create or find the best content, store it permanently, and receive dividends when the archive is viewed in the future. Through the use of the Koi Network and Arweave's Permaweb [Williams, 2018], participants will soon be able to earn lifelong dividends for submitting content that receives attention in the future. Further, the royalty streams created by these submissions to Koi are themselves tradable on Profit Sharing Token exchanges (e.g. Verto.exchange) — creating a liquid futures market for attention.

Anyone with an internet connection can install the Koi desktop app to act as a witness to this process, and receive a share of the rewards for helping to ensure that the archive is a valid and accurate representation of the original content. Additionally, we provide an aggregation layer for decentralized data marketplaces such as Ocean [McConaghy, 2020] to ensure that this archive can be collated by any member of the community and sold as indexed data for interested parties.

By solving the problem of web content incentivisation along a new axis — rewards proportional to attention — a new model can be created for the web. No longer are intrusive ads, sale of personal data, or freemium models of access required. In the new web, all information will be available to all people, with creators and builders rewarded for contributing data that other people find valuable — just as it should be.

For more information, please visit openkoi.com, or send a suggestion to hello@openkoi.com.

Contents

1 Incentivising Data Collation in the New Web

Koi solves a number of the traditional web's core problems— from disappearing links to accountability for 'fake' news and disinformation— by incentivizing the creation of an open knowledge archive on Arweave's permaweb. This archive is accessible by all, forever. The Koi network applies free-market principles with deliberate mechanism designs in pursuit of robust solutions to these issues. In this section, we discuss the motivations and details of these mechanisms.

Contributions to the Koi Network come in two forms: original creations and verified archives of content from the traditional web. In this section we describe the common reward mechanism for both of these contribution methods, as well as the validation protocol for verified archiving.

1.1 Mechanism Design

The Koi Network creates a content curation game, with participants being rewarded proportionately to the value of the data they create. Every piece of content submitted is permanently backed by Arweave's storage endowment (see section 2.1 for details). Once stored, content continues to indefinitely produce rewards when accessed.
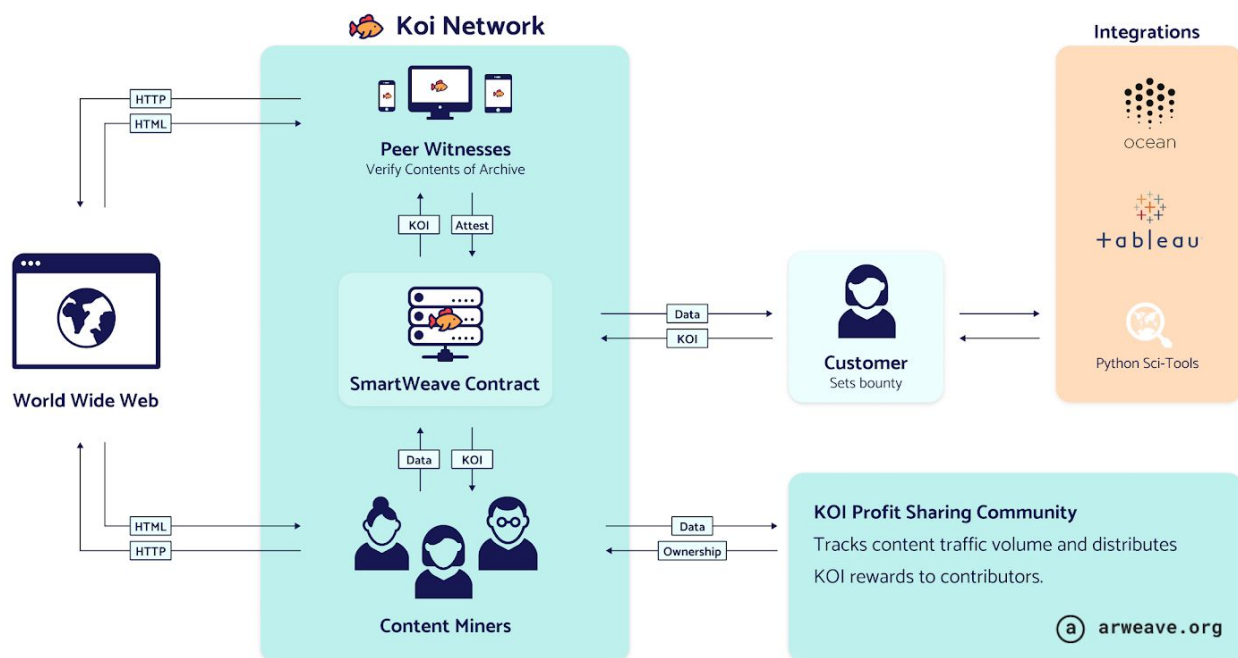

Figure 1: The Koi Network gathers data from original sources and across the traditional web, curating a high-quality permanent archive.

1.1.1 Game Phase Overview

The Koi game takes place across three continual phases: storage, optional verification, and reward distribution. Koi tokens are purchased and burnt in the storage phase, and new tokens are minted in the verification and distribution phases. This process creates an effective generation and sink lifecycle for tokens in the network.

1.1.2 Phase: Storage

During the first phase, **Content Miners** either create original content, or crawl the web for useful existing information. Once located, the content miner uploads this data to the Arweave's permaweb. After receiving a transaction identifier (TXID) for the newly stored content, the miner creates a new profit sharing token (PST) on the SmartWeave protocol. The TXID and PST are then bundled and submitted to the Koi profit sharing community (PSC) contract, along with a quantity of Koi tokens to burn.

1.1.3 Phase: Verification

Once stored, (non-original) content archived from the traditional web is verified by a network of **Peer Witness** nodes. During setup, peer witness nodes stake Koi tokens in order to ensure Sybil resistance and give veracity to their content attestations. In return, peer witnesses with higher stakes receive a greater proportion of verification rewards.

During operation, peer nodes monitor the Koi profit sharing community contract for newly added content. When archived (non-original) content is added, the peer witness automatically downloads the URL of the content on the centralised web, as well as the archived content on the permaweb. The sources are then compared and given a similarity score. This similarity score is signed by the peer witness and dispatched to the Koi PSC contract.

At the end of each reward period (initially expected to be 24 hours) the total number of attestations is calculated, and rewards distributed proportionately to contribution and stake. Specifically, the reward for each individual peer witness during a mining period is calculated as follows:

```
PeerWitnessReward = (sum( forall VerifiedArchives { if
                          PeerSubmittedAttestation { PeerStake /
                          TotalStakeOfAttestors } else { 0 } }) /
                          SumScoreForAllPeerWitnesses ) * RewardPeriodKOI
```

As a result of this mechanism, every piece of archived content from the traditional web is also paired with a weighted score, detailing its likelihood of being a legitimate copy of the original information.

1.1.4 Phase: Attention Reward Distribution

The final phase of the Koi game is the distribution of rewards proportionately to the attention that they have gained during each reward period. At the end of each reward period, Koi nodes collate anonymised traffic logs from the list of sources that the Koi community (via their contract) have presently decided are honest gateways to the permaweb. A small KOI reward is distributed to the runners of these gateways for the provision of logs.

After the collation process is complete, Koi nodes run fraud detection code across the provided logs, removing most instances of 'click fraud'. The fraud detection system employed at any given time is chosen by a periodic vote in the Koi community contract. This system of continual updates to the fraud prevention mechanisms of Koi allows it to adapt to new theft patterns in a similar way to the centralised approaches taken by projects like SteemIt, dTube, and Brave[2, 9].

Once logs have been collated and fraud attempts removed, reward distributions are calculated as follows:

```
distributeAttentionRewards(blockHeight, auditedTrafficLog)
    totalPeriodReward = calculatePeriodReward(blockHeight)
    weightedTotalTraffic = 0
    foreach TXID in auditedTrafficLog
        weightedTotalTraffic += getTokensBurntFor(TXID)
    foreach TXID in auditedTrafficLog
        (PST, burnQty) = getInfo(item)
        rewardAddr = selectPSTRewardAddr(PST)
        sendTokens(rewardAddr, burnQty / weightedTotalTraffic)
```

As well as rewarding miners proportionately to the popularity of their stored content in the future, this mechanism also encourages miners to compete to burn as many tokens as possible in the storage phase.

Further, this mechanism also allows the miner to trade 'futures' (via PSTs) on the profit that will be produced by their archived items in the future. These tokens will be freely tradable on any PST exchange (for example, Verto [4]). Automated bots will then be built to buy and sell portfolios of these PSTs to optimise them for likely future reward. In the process, these bots will generate early liquidity for content miners of the mechanism.

1.2 Additional Reward Modifiers

In addition to the outlined reward mechanisms above, modifications will be made at the discretion of the community to support the following initiatives:

1.2.1 Bounties

Production of curated, verified data sets is a valuable service that the Koi network can trivially offer to users, businesses, and enterprises. In order to capitalise on this, the Koi network can offer a service of web scraping

bounties. Customers of this service can provide a number of Koi tokens and a list of URLs that they would like to have archived. A portion of these tokens is then distributed to the owners of the Koi profit sharing community tokens, while the remainder is offered as immediate rewards for archivers and attestors that service the request.

1.2.1 Content Access Referral Rewards

In order to reward other permaweb services for encouraging their content creating users to buy, burn, and submit their data items to Koi, referrers given by URL parameters could be rewarded. For example, the [ArDrive](#) project intends to integrate Koi in order to further incentivise their users to upload data to the service. As an additional incentive for ArDrive, when ArDrive embeds content that has been submitted to the Koi network it uses a URL containing a reference to the ArDrive profit sharing community token contract. During the distribution phase, Koi would then distribute some of the newly minted tokens to the ArDrive community. Through this mechanism, the Koi community can expect a rapid adoption of their protocol across the ecosystem of permaweb apps.

1.3 Participant Roles

1.3.1 Peer Witnesses

In order to verify that the content submitted by miners is accurate and matches the original content of a web resource as it existed at time of archival, peer devices are incentivized to verify and attest to the validity of the permaweb archive. This accomplishes two things: first, it ensures a transparent chain of accountability for the archived data, and second, it removes KOI from circulation in the form of staking, and distributes rewards to any peer that participates. Peer witnesses do not need to be technologically sophisticated, and can simply install a node client to begin participating and earn KOI.

1.3.2 Content Miners

Web crawling is a mature technology in which a programmatic agent navigates through the internet from page to page, seeking content. While this is the basis of a majority of online content gathering systems (search engines, analytics tools, and more) there is currently no open marketplace for these businesses to compete. Koi provides incentives for these existing players to expose some portion of their data to the public and to contribute to our archive.

1.3.3 Data Curators / Customers

Whether for personal use, or on a larger scale as a business, a wide range of use cases exist for gathering online data. A customer of the Koi Network can easily set a bounty to request certain data types be collected from the web. Miners are thus incentivized to focus on those areas to offset the costs of their web crawling efforts.

2 Technology

The Koi stack combines mature application-layer technologies such as HTML parsing and headless browsing with more modern tools like the Arweave permaweb and token-based economics ("tokenomics") to provide a solution that incentivises the curation of a high quality, open archive.

2.1 The Arweave's Permaweb

The Arweave network has established a permanent archival web, known as the permaweb [Williams, 2018]. The Arweave protocol is similar to Bitcoin [Nakomoto, 2008] or Ethereum [Buterin, 2015] in structure, but makes use of a more scalable storage architecture, as well as sustainable economic endowments to back permanent replication of data. By employing a Succinct Proof of Random Access mining mechanism, nodes in the Arweave network are incentivised to make many hundreds of thousands of replications of the data added to the network.

2.1.1 Profit-Sharing Communities

Profit sharing communities like Koi are a new model of startups, created for building and governing decentralised web applications and services [Community.xyz, 2020]. Rather than accruing value in a centralised treasury, profit sharing communities typically issue this profit as dividends to those that have built, contributed to, or invested in the service through the form of a profit sharing token (PST).

2.1.2 Profit-Sharing Tokens

Profit sharing tokens (PSTs) are shares of royalty streams. These tokens — represented on the permaweb's smart contracting platform (SmartWeave) — can be held in order to accrue value from the stream, or traded using a PST exchange (for example, Verto). The Koi PSC contract has a profit sharing token, as well as each individual piece of content. This allows traders to hold the PSC token in order to accrue tips from interactions with the protocol, as well as to trade interests in individual and grouped archived items.

2.1.2 Attention Tracking

Previous blockchain projects such as Steemit [Steemit, 2018] and the Basic Attention Token [Brave, 2018] have established the feasibility of tracking 'reads' of information through nodes of the network. On the Arweave Permaweb, gateways — such as those in the Amplify network — offer anonymised access logs which can be used to track the popularity of archived items inside the permaweb.

2.1.3 Storage Throughput

Unlike typical proof-of-work blockchains, Arweave utilizes a 'blockweave' that can be expanded asynchronously, and thereby supports a much higher rate of content creation, and has much lower per-transaction costs. This technology is a necessity for the volume of content that Koi will create, but also for the feasibility of micropayments, which is not economically viable on Ethereum or traditional proof-of-work blockchains.

2.2 Koi Network

The Koi Network exists to aid the curation of an open, permanent archive of highly valuable information. This process follows three key steps: **Encode**, **Crawl**, and **Verify**; see more details below.

2.2.1 Web Encoding

As a matter of necessity for archived data from the traditional web, the Koi Network creates a structured representation of each page added and maps the resource to identify the useful parts of the data. Since only the important parts of the page are normally archived, the network is able to reach consensus on complex content pages, and peer witnesses can easily verify that the content matches the archive.

When a bounty is set, either by the DAO or a customer, specific parts of the web page are identified for archival purposes. Miners have the freedom to archive additional information if they believe it may become valuable in the future, earning rewards for it through the issued PST.

2.2.2 Profitable Web Crawling

Many organizations already engage in web crawling. The Koi Network gives them a way to monetize their findings by receiving dividends for submitting content to the permaweb. In addition, crawlers can prioritize URLs which have existing KOI bounties, and thereby offset the cost of their scraping even further. Koi will provide easily implemented hooks to support integration by any existing parties, as well as easy to use templates which can be deployed by hobbyists on consumer hardware.

2.2.3 Low-Cost 'Witness Mining'

Because peer nodes are not storing any data or performing costly operations like hashing, the Koi witness node can operate on a wide range of devices with minimal cost to the end user. As such, Koi provides an easy to access revenue stream for anyone with an internet connection, as well as the opportunity to participate in the creation and governance of a truly decentralized human knowledge archive.

3 Tokenomics

The KOI token is used to interact with the Koi Network, and is designed to reward archive contributors over the long-run. Most importantly, the day-to-day function of the network will reduce circulating supply through token burning and staking of KOI tokens. The core value of the network as a curator of valuable data will ensure demand for the tokens as they are required to access these services.

3.1 Applications & Utility

In 2021, the market for machine learning training data is expected to top USD $10B [McKinsey, 2017], and a wide range of businesses from e-commerce companies to airlines use online data to drive decision-making. The following are some examples of the range of utility provided by web data, and how the Koi Network will commoditize and monetize it.:

- Content archives (historical reference data is stored on Arweave and other archives)
- Machine learning training data (provides images and text, labelled consistently)
- Public record data in machine-friendly format (e.g. licensing or financial data)
- Product catalogues, comparative pricing, and general e-commerce data
- Market data (flight prices, commodity information, weather, etc.)

As the network grows, the DAO will provide grants and funding to integrate KOI as a data mining tool with existing data analysis suites such as Python Scitools and Tableau, so that customers can easily take advantage of our network of on-demand web scraping tools and nodes.

3.2 Long Term Value Creation

As the network grows, the main driver of token value will be the desire to add to the permanent archive and receive dividends. Anyone can upload content by burning KOI, which creates an open contest among creators, archive bots, and anyone else with an internet connection. This activity results in valuable data assets that provide long term dividends, enabling a new form of wealth creation by helping to preserve history and curate a deeply valuable archive.

3.2.1 Implicit Content Appraisal

By creating an open marketplace for content archiving, we also establish the future value of current content. This will provide useful information for content moderators or those seeking to build search engines and other technologies that provide content hierarchies.

3.2.2 Indexing and Search

By keeping the attention data of this network public, Koi will create a new paradigm under which any internet user can deploy a personal intelligent agent to search on their behalf, without the need to consult a formal authority to retrieve web information.

3.2.3 Data Marketplaces

The notion of data marketplaces is quite old, but they have become increasingly popular in part due to the recent surge in demand for training data for artificial intelligence and corporate decision making. Anyone is welcome to repackage the Koi data sets, process them, clean them, and sell them through data marketplaces (for example, Ocean Protocol). Ultimately, this practice will further the consensus of the network and add to the attention calculus, leading to a stronger overall network and a more tightly integrated permaweb.

3.2.4 Single-Version Data Resources

It is beginning to become clear that web resources have a real world cost, not only to the bottom line, but to the environment as well. By establishing a standard for singular storage of content archives on reliable long term systems such as the permaweb, Koi reduces substantially the need for vast cloud computing resources, and can in fact help businesses to streamline their archival operations, while also ensuring authenticity and permanence.

3.2.5 Corporate Communications

As the financial sector is increasingly driven by short term speculation, there has been a growing demand for accountability in public relations statements and press releases. In particular, businesses have begun to explore the use of blockchain technology as a means of securing their investors against market manipulation [Verizon, 2020]. Koi elegantly provides a solution for these customers by ensuring a ready group of witnesses for any such statements, as well as certainty of archival stability, without the typical overhead costs.

References

1. Berners-Lee, T. Semantic Web Road map. http://www.w3.org/DesignIssues/Semantic.html (last modified Oct. 14, 1998)

2. Brave Software, "Basic Attention Token", https://basicattentiontoken.org/wp-content/uploads/2017/05/BasicAttentionTokenWhitePaper-4.pdf, 2018

3. Community.xyz Documentation, Arweave, https://github.com/CommunityXYZ/community-js/blob/master/docs/classes/_community_.community.md, 2020

4. Marshall, C.C. and Shipman, F.M. n.d. "Which Semantic Web." [Online]. Available WWW: http://www.csdl.tamu.edu/~marshall/ht03-sw-4.pdf. 2008

5. Mary L. Gray and Siddharth Suri, "Ghost Work" 2019

6. McKinsey & Company, "Artificial Intelligence, the Next Digital Frontier", https://www.mckinsey.com/~/media/mckinsey/industries/advanced%20electronics/our%20insights/how%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/mgi-artificial-intelligence-discussion-paper.ashx, 2017

7. Salih Ismail1 and Talal Shaikh2, "A Literature Review on Semantic Web - Understanding the Pioneers' Perspective" Sixth International Conference on Computer Science, Engineering & Applications DOI: 10.5121/csit.2016.61102 05 Sep 2016

8. S. Williams, V. Diordiiev, L. Berman, I. Raybould, I. Uemlianin, "Arweave: A Protocol for Economically Sustainable Information Permanence" https://www.arweave.org/yellow-paper.pdf, 2018

9. Steemit, "Steem", https://steem.com/steem-whitepaper.pdf, 2018

10. Trent McConaghy, "Ocean Protocol: Tools for the Web3 Data Economy", https://oceanprotocol.com/tech-whitepaper.pdf 2020

11. Verizon "Full Transparency" https://www.verizon.com/about/news/verizons-full-transparency-launches-blockchain-verification, 2020

12. Yuji Roh, Geon Heo, Steven Euijong Whang, Senior Member, IEEE, "A Survey on Data Collection for Machine Learning" arXiv:1811.03402v2 [cs.LG] 12 Aug 2019

If this project interests you, [run a node](#) now and get rewarded for being an early adopter.

Earn KOI, and help us preserve human knowledge.

You can make a difference.