

# Ocean Protocol:

## Tools for the Web3 Data Economy

Ocean Protocol Foundation<sup>1</sup> with BigchainDB GmbH<sup>2</sup>

Technical Whitepaper, Version 2022-SEP-01

### Abstract

Ocean Protocol is an on-ramp for data services into crypto ecosystems, using *data NFTs* and *datatokens*. Data NFTs are like master tapes, and datatokens are like CDs: base IP and licenses respectively. A data NFT is a non-fungible ERC721 token representing copyright of a data service; a datatoken is a fungible ERC20 token to access the service. Ocean smart contracts and libraries make it easy to publish data services (deploy and mint data NFTs & datatokens) and consume data services (spend datatokens). Ocean contracts are deployed to Ethereum mainnet and other EVM networks. Ethereum composability enables crypto wallets as data wallets, crypto exchanges as data marketplaces, DAOs as data coops, datatoken-backed stable assets, and more.

Ocean Market is an open-source community marketplace for data.

Ocean's "Compute-to-Data" feature gives compute access on *privately held* data, which never leaves the data owner's premises. Ocean-based marketplaces enable monetization of private data while preserving privacy.

Token dynamics are designed such that OCEAN health rises with usage volume.

The community-driven OceanDAO funds software development, outreach, and more. veOCEAN is a means to lock OCEAN to be able to vote in OceanDAO and earn ecosystem rewards. veOCEAN holders can earn passively, but can earn more by actively allocating veOCEAN to data assets that have high data consume volume (DCV).

We envision thousands of data marketplaces, where Ocean Market is just one. In addition to Ocean Market being open-source (and therefore forkable), Ocean includes tools to help developers build their own marketplaces and other apps.

These tools are part of a system designed for long-term growth of a permissionless Web3 Data Economy. The Ocean Data Farming program incentivizes a supply of data.

---

<sup>1</sup> [oceanprotocol.com](https://oceanprotocol.com)

<sup>2</sup> [bigchaindb.com](https://bigchaindb.com)

# Contents

<b>1. Overview</b>	<b>4</b>
1.1. Introduction	4
1.2. Paper Organization	4
<b>2. Top-Level Block: Ocean System</b>	<b>5</b>
2.1. Goals	5
2.2. The Design	5
2.3. Kickstarting Growth, and Long-Term Sustainability	5
<b>3. Subblock: Data Ecosystem Powered by Ocean Tools</b>	<b>6</b>
3.1. Introduction	6
3.2. USPs of Ocean Tools	6
3.3. Ocean Tools Foundation: Data NFTs & Datatokens	7
3.3.1. Introduction	7
3.3.2. Mental Model	7
3.3.3. IP Backgrounder	7
3.3.4. IP <> Data NFTs & Datatokens	8
3.3.5. Datatokens are ERC20 Access Tokens	8
3.3.6. Data Access Variants	9
3.4. Ocean Tools Architecture	10
3.4.1. Overview	10
3.4.2. Data NFTs, Datatokens, and Access Control Tools	10
3.4.3. Market Tools	10
3.4.4. Metadata Tools	11
3.4.5. Third-Party Apps & Tools	11
3.4.6. Actor Identities	11
3.5. Ocean Tools: Multi-Network Deployment	11
3.6. Ocean Tools: Data Marketplaces	12
3.6.1. Introduction	12
3.6.2. Marketplaces Architecture	12
3.6.3. Ocean Marketplace Actions	13
3.6.4. Developer Tools & 3rd Party Marketplaces	13
3.6.5. Sustainability & Fees	14
3.6.6. IP Violations and Sensitive Data	14
3.6.7. Benefits of Ocean Data Marketplaces	15
3.7. Ocean Tools: Compute-to-Data	16
3.7.1. Motivation	16
3.7.2. Conceptual Working	17

3.7.3. Compute-to-Data Flow Variants	17
3.7.4. Share, or Monetize	18
3.7.5. Compute-to-Data Architecture	18
3.7.6. Marketplaces and Compute-to-Data	18
3.7.7. Trusting Algorithms	19
3.7.8. Benefits of Compute-to-Data	19
3.7.9. Relation to other Privacy-Preserving Technologies	19
3.8. Ocean Tools: Fine-Grained Permissions	19
3.8.1. Introduction	19
3.8.2. Market-Level Permissions	20
3.8.3. Asset-Level Permissions	20
<b>4. Subblock: OceanDAO</b>	<b>21</b>
4.1. Introduction	21
4.2. OceanDAO Components	21
4.3. OceanDAO Implementation	21
<b>5. Ocean Data Farming</b>	<b>22</b>
5.1. Overview	22
5.2. Reward Function	22
<b>6. On OCEAN Token Model</b>	<b>23</b>
6.1. Introduction	23
6.2. Token Value $\propto$ Volume	23
6.3. Work Token (and Governance Token)	23
6.4. Burning	24
6.5. veOCEAN	24
6.6. User Acquisition	24
6.7. On OCEAN as Unit-of-Exchange	24
<b>7. Ocean Applications</b>	<b>25</b>
7.1. Decentralized Orchestration	25
7.2. Data Wallets: Data Custody & Data Management	25
7.3. Data Auditability	26
7.4. Data DAOs: Data Co-ops and More	26
7.5. Fractional Ownership	26
7.6. Group-Restricted Data Access	27
7.7. Unlocking Latent Data of Individuals & Enterprises	27
7.8. Data Marketplaces	28
7.9. Data as an Asset Class for DeFi	29
7.10. Data to Optimize DeFi Returns	29

7.11. Data Management Platforms for Smart Cities and More	30
7.12. Composable Data NFTs & Datatokens	30
<b>8. Conclusion</b>	<b>30</b>
<b>9. References</b>	<b>31</b>

# 1. Overview

## 1.1. Introduction

Modern society runs on data [\[Economist2017\]](#). Modern artificial intelligence (AI) extracts value from data. More data means more accurate AI models [\[Banko2001\]](#) [\[Halevy2009\]](#), which in turn means more benefits to society and business. The greatest beneficiaries are companies that have *both* vast data and internal AI expertise, like Google and Facebook. In contrast, AI startups have excellent AI expertise but are starving for data; and typical enterprises are drowning in data but have less AI expertise. The power of both data and AI—and therefore society—is in the hands of few.

The aim of Ocean Protocol is to spread the benefits of AI by equalizing the opportunity to access and monetize data. We accomplish this by creating simple *tools* to publish data and consume data as decentralized *data NFTs & datatokens*. Data NFTs & tokens interoperate with ERC721 & ERC20 wallets, exchanges, DAOs and more. This data may be held on-premise to preserve privacy. Additionally, Ocean has tools for data marketplaces. We have implemented these tools as Solidity code running on Ethereum mainnet [\[Buterin2013\]](#) and other EVM networks [\[EthereumEvm2022\]](#); as Python and JavaScript libraries to ease higher-level integration; and as a community data marketplace web application.

These tools are encapsulated in a broader system design for long-term growth of an open, permissionless data economy. The Data Farming program incentivizes a supply of data. OceanDAO funds software development, outreach, and more. OceanDAO will be funded by revenue from apps and services in the Ocean data ecosystem, Ocean network rewards, and Ocean Protocol Foundation. These are all detailed in the paper.

Everything described above has been deployed live as of Nov 30, 2020.

## 1.2. Paper Organization

The rest of this paper is organized as follows.

**Ocean System.** Section 2 describes Ocean’s system-level design, which is designed to facilitate the emergence of an open data ecosystem.

**Ocean Subblocks.** Then, several sections describe key building blocks in the Ocean system:

- Section 3 describes Ocean Tools to power a Data Ecosystem, including data NFTs & datatokens, tools software architecture, network deployment, data marketplaces, and Compute-to-Data.
- Section 4 describes OceanDAO.

**Ocean Data Farming.** Section 5 describes a program to kickstart a supply of quality data.

**OCEAN token.** Section 6 describes the OCEAN token model, bringing together information presented in prior sections.

**Applications.** Section 7 describes applications of Ocean.

**Final.** Section 8 concludes. Section 9 holds references.

## 2. Top-Level Block: Ocean System

### 2.1. Goals

The top-level goal is to spread the benefits of AI by equalizing the opportunity to access and monetize data. We can make this more specific. The Ocean System has these aims:

- An overall system that is **sustainable and growing**, towards ubiquity.
- Basic design is **simple** to understand and communicate.
- In line with **Ocean Mission and Values** [McConaghy2018]. These include: unlock data, preserve privacy, spread of power, spread of wealth, stay within the law, censorship-resistant & trustless. It should be permissionless, rent-free, and useful to the world [Goldin2017]. It should be anti-fragile: get more resilient when “kicked”, therefore also needing evolvability. It follows time scales of decades not months.

### 2.2. The Design

Figure 1 shows the Ocean System design. At its heart is a **loop**, designed for “snowball effect” growth of the ecosystem. The **Workers** (center) do **\*work\*** to help grow the **Data Ecosystem** (right). Marketplaces and other data ecosystem services generate revenue, using Ocean software tools. A tiny fraction of that revenue is **looped back** (arrow looping from right to left) as **Network revenue** to the Ocean community: to **Buy & Burn OCEAN** (bottom left) and back to workers curated by **OceanDAO** (center-left). To catalyze growth and ensure decent funding in early days, **Network rewards** (left) also feed to Workers via OceanDAO.

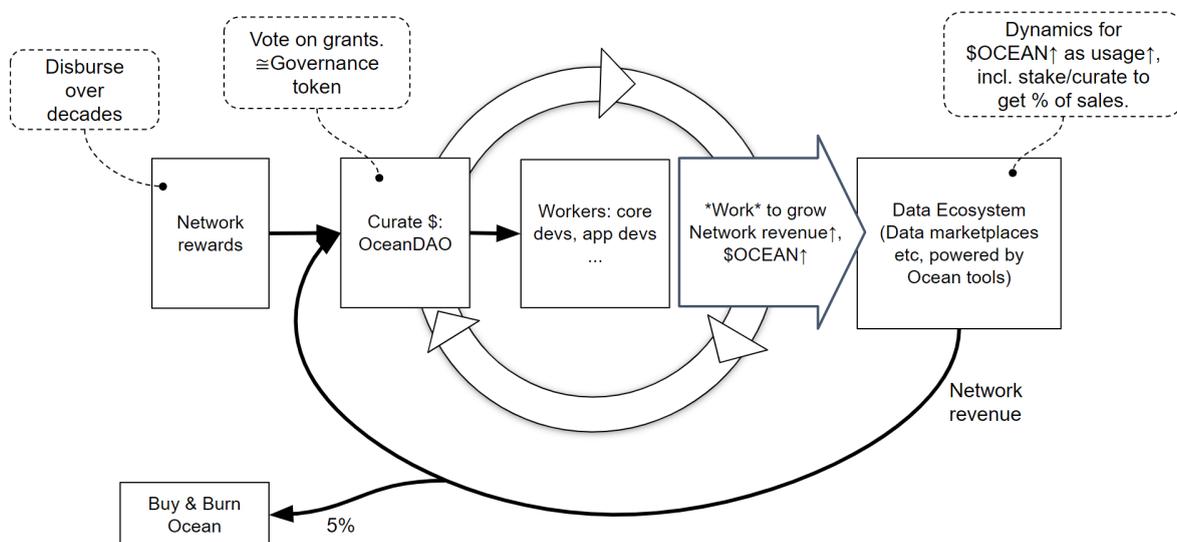


Figure 1: Ocean System

### 2.3. Kickstarting Growth, and Long-Term Sustainability

The system design achieves sustainability over decades via funding that goes to teams (Workers) that continually add value to the Data Ecosystem.

Funding comes from Network Revenue and from Network Rewards. In the longer term, the bulk of revenue will be from Network Revenue. To help kickstart growth, **51%** of the overall OCEAN token

supply is dedicated to Network Rewards, which follow a Bitcoin-like disbursement schedule (but with a ramp-up period).

**The Ocean token is designed such that as usage of Ocean grows, it grows the value accrued in OCEAN.** Here's the loop: more usage of Ocean tools (in the data ecosystem) leads to more Network Revenue, which goes to burning and OceanDAO. Burning OCEAN reduces supply, to grow the value accrued in OCEAN. Funds go through OceanDAO to workers who have the mandate to grow usage of Ocean tools. And the loop repeats.

OceanDAO curates funding coming from Network Revenue and Network Rewards. It gets people to do "work" such as improve Ocean core software, build applications using Ocean, spread awareness, grow data supply, and more. A Data Farming program incentivizes data supply.

This system design generalizes beyond Ocean; we can call it the "Web3 Sustainability Loop". It draws inspiration from businesses and nations. The essay [\[McConaghy2020d\]](#) elaborates.

### 3. Subblock: Data Ecosystem Powered by Ocean Tools

This section and the ones that follow add details on subblocks of the Ocean system.

#### 3.1. Introduction

The Data Ecosystem is a major subblock of the Ocean system. **Ocean tools** power the Data Ecosystem. This section describes Ocean tools.

#### 3.2. USPs of Ocean Tools

Towards *building something that people want* [\[Graham2005\]](#), Ocean tools offer these unique selling propositions (USPs):

- **Earn by selling data, and staking on data.** Ocean Market makes it easy to sell your data, whether you are an individual, company, or city. Furthermore, users can stake on data: lock OCEAN then allocate veOCEAN to datasets for rewards.
- **An on-ramp and off-ramp for data assets into crypto**, allowing: crypto wallets for data custody & data management, DEXes<sup>3</sup> for data exchanges, DAOs<sup>4</sup> for data co-ops, datatoken-backed stable assets like H2O [\[NewOrder2021\]](#), securitizing data assets, and more via DeFi composability. It's "data legos." The data itself does not need to be on-chain, just the access control.
- **Quickly launch a data marketplace, with many USPs:** buy & sell private data while preserving privacy, non-custodial, censorship-resistant, auto price discovery, data audit trails, and more.
- **Decentralized data exchange platform**, enabling these characteristics: improve the visibility, transparency and flexibility in usage of data; share data while avoiding "data escapes"; needs little dev-ops support and maintenance; has high liveness; is non-custodial; and is censorship-resistant. E.g. Ocean as a traffic data management platform for smart cities. E.g. federated learning without having to trust the orchestration middleman [\[McConaghy2020c\]](#).

---

<sup>3</sup> DEX = Decentralized Exchange

<sup>4</sup> DAO = Decentralized Autonomous Organization

### 3.3. Ocean Tools Foundation: Data NFTs & Datatokens

#### 3.3.1. Introduction

Ocean is an on-ramp and off-ramp for data assets into crypto ecosystems, using *data NFTs* and *datatokens*. With analogy to music: Data NFTs are like master tapes, and datatokens are like CDs: base IP and licenses respectively.

- A data NFT is a non-fungible ERC721 token [\[ERC721\]](#) representing “base IP” - copyright or exclusive rights of a data service. This is the equivalent to a music “master tape”.
- A datatoken is a fungible ERC20 token [\[Vogelsteller2015\]](#); having 1.0 datatokens for a given data service means you can access that service; it’s a license against the base IP. This is the equivalent to a music CD.

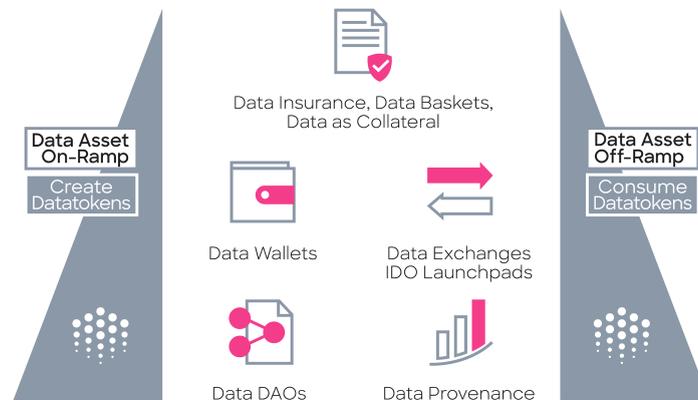
Each data service gets its own data NFT; and zero or more datatokens against that data NFT.

Ocean smart contracts and libraries make it easy to publish data services (deploy and mint data NFTs & datatokens) and consume data services (spend datatokens).

This setup helps keep Ocean as simple as possible (but no simpler), composable, and make Ocean a *set of tools* (versus a platform that might lock you in).

#### 3.3.2. Mental Model

Figure 2 shows the mental model. Ocean does the beginning (create data assets) and the end (consume data asset). In between are any ERC20-based applications, including Ocean-based marketplaces.



**Figure 2: Mental model. Data NFTs & datatokens are the interface to connect data assets with DeFi tools. Ocean is an on-ramp for data services into data NFT & datatoken data assets on Ethereum, and an off-ramp to consume data.**

#### 3.3.3. IP Backgrounder

Here’s a quick background on IP, via the example of music. It also applies for books, videos, and data. A more thorough treatment can be found in [\[McConaghy2021a\]](#).

**Copyright.** When Michael Jackson recorded “Billie Jean”, he automatically got copyright. Copyright laws for USA, Germany, and 177 other countries follow the 1983 Berne Convention [\[WIPO2022\]](#), which provides creators “with the means to control how their work is used, by whom, and on what terms”. In it, the “creator of a work doesn’t need to apply for copyright, or to register it.” Rather,

copyright automatically exists once it has been put it into “a fixed form”, e.g. as a recording on audio tape or as magnetic bits on a hard disk.

**Exclusive Rights.** When Universal Music bought the rights to Michael Jackson’s recording of the song “Billie Jean”, Universal got an exclusive license (a contract) to use that song for their own profit. The base IP was transferred to Universal.

**Sub-licenses.** When you buy a music CD, you’re buying the license to listen to it, for personal consumption. But you can’t resell it or play it on the radio, since you don’t have the license for that.

### 3.3.4. IP <> Data NFTs & Datatokens

Ocean uses both ERC721 data NFTs and ERC20 datatokens.

A **data NFT** is a non-fungible ERC721 token representing “base IP” - copyright or exclusive rights of a data service. Base IP rights include the ability to create access licenses (=deploy ERC20 datatokens, and mint them tokens). You can transfer a data NFT to another person, for them to hold those exclusive rights.

A **datatoken** is a fungible ERC20 token; having 1.0 datatokens for a given data service means you can access that service; it’s a license against the base IP. [\[McConaghy2021c\]](#) has details.

Note: Ocean V3 (released fall 2020) used just ERC20 datatokens. The base rightsholder was implicit: the publisher, who can mint more ERC20 tokens. [\[McConaghy2021b\]](#) has details. We moved to data NFTs and datatokens for Ocean V4 (released spring 2022).

Note: We could have used ERC721 for data access control, where you can access the dataset if you hold the token. However, datasets typically get shared among >1 people. For this we need fungibility, which is the realm of ERC20. ERC20 brings many more benefits, especially easy price discovery via various data pricing approaches.

### 3.3.5. Datatokens are ERC20 Access Tokens

Traditional access tokens exist, such as OAuth 2.0 [\[OAuth2020\]](#). If you present the token, you can get access to the service. However, the “tokens” are simply a string of characters, and “transfer” is basically copying and pasting that string. This means they can easily be “double-spent”: if one person gets access, they can share that access with innumerable others, even if that access was only meant for them. These tokens aren’t the “tokens” we think of in blockchain.

How do we address the double-spend problem? This is where blockchain technology comes in. In short, there’s a single shared global database that keeps track of who owns what, and can then easily prevent people from spending the same token twice.

This generalizes beyond Bitcoin tokens to other token assets. ERC20 [\[Vogelsteller2015\]](#) was developed as a standard for token ownership actions. It’s been adopted widely in Ethereum and beyond. Its focus is *fungible* tokens, where tokens are fully interchangeable.

We can connect the idea of *access* with the ERC20 token standard. Specifically, consider an ERC20 token where **you can access the dataset if you hold 1.0 tokens**. To access the dataset, you send 1.0 datatokens to the data provider. You have custody of the data if you have at least 1.0 tokens. To give access to someone else, send them 1.0 datatokens. That’s it! But now, the double-spend problem is

solved for “access control”, and by following a standard, there’s a whole ecosystem around it to support that standard.

### 3.3.6. Data Access Variants

At the smart contract level, data NFTs & datatokens don’t differ. Variants emerge in the semantic interpretation by libraries run by the data provider, one level up. Here are some variants:

- Access could be **perpetual** (access as many times as you like), **time-bound** (e.g. access for just one day, or within specific date range), or **one-time** (after you access, the token is burned).
- Data access is always treated as a data service. This could be a service to access a **static** dataset (e.g. a single file), a **dynamic** dataset (stream), or for a compute service (e.g. “bring compute to the data”). For static data, we can tune variants based on the type of storage: Web2 cloud (e.g. AWS S3), Web3 non-permanent (e.g. Filecoin), Web3 permanent small-scale (e.g. Ethereum), Web3 permanent large-scale (e.g. Arweave), or go meta using IPFS but “pinned” (served up) by many places. For dynamic data, variants include Web2 streaming APIs (single-source), Web3 public data oracles (e.g. Chainlink), and Web3 private data oracles (e.g. DECO).

The terms of access are specified in the metadata, which is on-chain (more on this later).

## Architecture

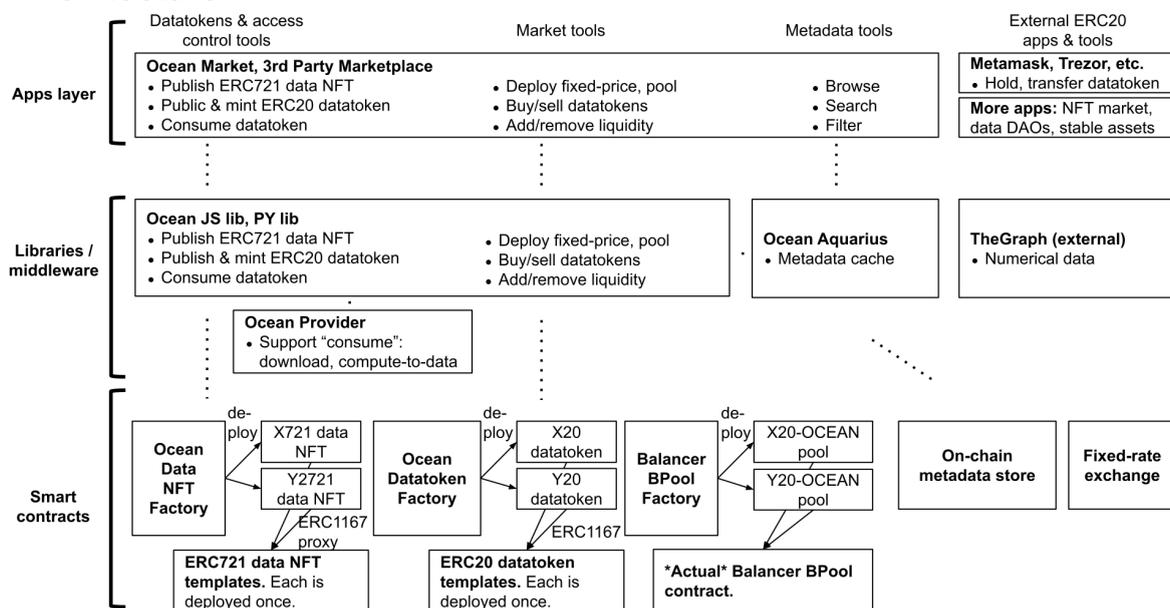


Figure 3: Ocean tools architecture

## 3.4. Ocean Tools Architecture

### 3.4.1. Overview

Figure 3 shows the Ocean tools architecture. The lowest level has the smart contracts, which are deployed on Ethereum mainnet and other networks. Above that are libraries and middleware, which expose the contracts to higher-level languages and provide convenience utilities. The top layer is applications.

Left to right are groupings of functionality: tools for data NFTs & datatokens, tools for markets (including pools), tools to consume data services and for metadata, and external ERC20 tools. The following subsections elaborate; more information yet is in the documentation [OceanDocs2022] and the open-source code itself [OceanContracts2022] [OceanLibPy2022] [OceanLibJs2022][OceanMarket2022].

### 3.4.2. Data NFTs, Datatokens, and Access Control Tools

The publisher actor holds the dataset in Google Drive, Dropbox, AWS S3, on their phone, on their home server, etc. The dataset has a URL. The publisher can optionally use IPFS for a content-addressable URL. Or instead of a file, the publisher may run a compute-to-data service.

In the **publish** step, the publisher invokes **Ocean Data NFT Factory** to deploy a new data NFT to the chain. The publisher can then invoke **Ocean Datatoken Factory** to deploy a new datatoken contract (against the data NFT) to the chain), then mints datatokens. To save gas fees, deployments use ERC1167 proxy approach on templates for the data NFT and datatoken.

The publisher runs **Ocean Provider**. In the **consume** step, Provider software needs to retrieve the data service URL given a datatoken address. One approach would be for the publisher to run a database; however this adds another dependency. To avoid this, it stores the URL on-chain. So that others don't see that URL, it encrypts it.

To initiate the **consume** step, the data consumer sends 1.0 datatokens to the Provider wallet. Then they make a service request to the Provider. The Provider loads the encrypted URL, decrypts it, and provisions the requested service (send static data, or enable a compute-to-data job). The Appendix has details of this process.

Instead of running a Provider themselves, the publisher can have a 3rd party like Ocean Market run it. While more convenient, it means that the 3rd party has custody of the private encryption/decryption key (more centralized). Ocean will support more service types and url custody options in the future.

**Ocean JavaScript and Python libraries** act as drivers for the lower-level contracts. Each library integrates with Ocean Provider to provision & consume data services, and Ocean Aquarius for metadata.

### 3.4.3. Market Tools

Once someone has generated datatokens, they can be used in any ERC20 exchange. In addition, Ocean provides **Ocean Market**, a reference data marketplace for use by the Ocean community. It's decentralized (no single owner or controller), and non-custodial (only the data owner holds the keys for the datatokens).

Ocean Market supports fixed pricing. For fixed pricing, there's a simple contract for users to buy/sell datatokens for OCEAN, while avoiding custodianship during value transfer.

Complementary to Ocean Market, Ocean has reference code to ease building **third-party data marketplaces**, such as for mobility (Daimler [\[OceanBlog2020b\]](#)).

#### 3.4.4. Metadata Tools

Metadata (name of dataset, dateCreated etc.) is used by marketplaces for data asset discovery. Each data asset can have a decentralized identifier (DID) [\[DID2019\]](#) that resolves to a DID document (DDO) for associated metadata. The DDO is essentially JSON filling in metadata fields.

Ocean uses the blockchain network as an **on-chain metadata store**, i.e. to store both DID and DDO. This means that once the write fee is paid, there are no further expenses or dev-ops work needed to ensure metadata availability into the future, aiding in the discoverability of data assets. It also simplifies integration with the rest of the Ocean system, which is Ethereum-based. Storage cost on blockchain networks is not negligible, but not prohibitive and the other benefits are currently worth the tradeoff compared to alternatives.

Due to the permissionless, decentralized nature of data on Ethereum mainnet (and other EVM networks), any last-mile tool can access metadata. **Ocean Aquarius** supports different metadata fields for each different Ocean-based marketplace. Third-party tool **TheGraph** sees metadata fields that are common across all marketplaces.

To address GDPR right-to-be-forgotten, metadata is stored *encrypted*. The publisher running a Provider provides a REST service to get decrypted metadata. Typically, an Ocean-powered market will retrieve decrypted metadata from various publishers, then store the decrypted metadata in its local Aquarius.

#### 3.4.5. Third-Party Apps & Tools

Because they are ERC721 & ERC20, data NFTs & datatokens have great potential composability with other Ethereum tools and apps. This includes including **MetaMask** and **Trezor** as data wallets, DEXes as data exchanges, and more. The Applications section expands on this.

#### 3.4.6. Actor Identities

Actors like data providers and consumers have Ethereum addresses, aka web3 accounts. These are managed by crypto wallets, as one would expect. For most use cases, this is all that's needed.

In Ocean Market, publishers can opt-in to provide more of their identity information via 3Box [\[3Box2020\]](#), to help data buyers build their trust in the publisher. To go further, the Ocean community could layer on protocols like Verifiable Credentials [\[W3C2019\]](#).

This subsection has described the Ocean tools architecture at a higher level. [\[McConaghy2020h\]](#) [\[OceanDocs2022\]](#) has further details.

### 3.5. Ocean Tools: Multi-Network Deployment

Towards a broadly open Data Economy, we aim for Ocean deployment across many chains as a thin layer for data assets and permissioning.

- **Ocean is deployed to Ethereum mainnet and other EVM networks** (for lower gas costs) including Polygon, Binance Smart Chain (BSC), Moonriver, and Energy Web Chain (EWC). A full list is at [\[OceanDocsNetworks2022\]](#).
- **The contracts do *not* have upgradeability built in.** Therefore the only way to upgrade contracts is by *community consensus* to use a new set of smart contracts. This principle of

governance minimization shrinks the attack surface, and also avoids security issues specific to upgradeable smart contracts.

Over time, Ocean will get deployed to more L1 EVM networks, L2 rollups, and eventually, non-EVM networks. On a longer time horizon, we envision *every* blockchain network to have an Ocean-powered data resource permissioning layer. We envision them being interconnected with data NFTs, datatokens and OCEAN flowing everywhere.

In all these deployments, the root OCEAN token deployment will remain on Ethereum mainnet, and bridged to other networks.

### 3.6. Ocean Tools: Data Marketplaces

#### 3.6.1. Introduction

This section drills into details about Ocean-based data marketplaces.

Since each data service has its own ERC20 token, any ERC20 exchange can serve as a data marketplace. We can still make it easier for users. Specifically, marketplaces *tuned* for data can help users in the whole data flow, including publish data, set price, curate data, discover data, buy data, and consume data.

#### 3.6.2. Marketplaces Architecture

Figure 4 shows the conceptual architecture. There are many data marketplace frontends; there is a common backend (for a given network).

The frontends include a community market (**Ocean Market**) [top middle] and independent third-party markets like those of Daimler [top left]. Each frontend runs client-side in the browser, using **Ocean JS library**, which interfaces with the backend.

The **Decentralized Backend** [bottom left] is Solidity code running on the EVM chain, and includes the data NFT & datatoken and pool contracts, and the on-chain metadata store.

When a **Buyer** purchases a dataset on a frontend [far top, far left], most of the revenue goes to the **Data seller** [bottom right]. Some fees go to **LPs, the marketplace runner, and the broader community** [top right].

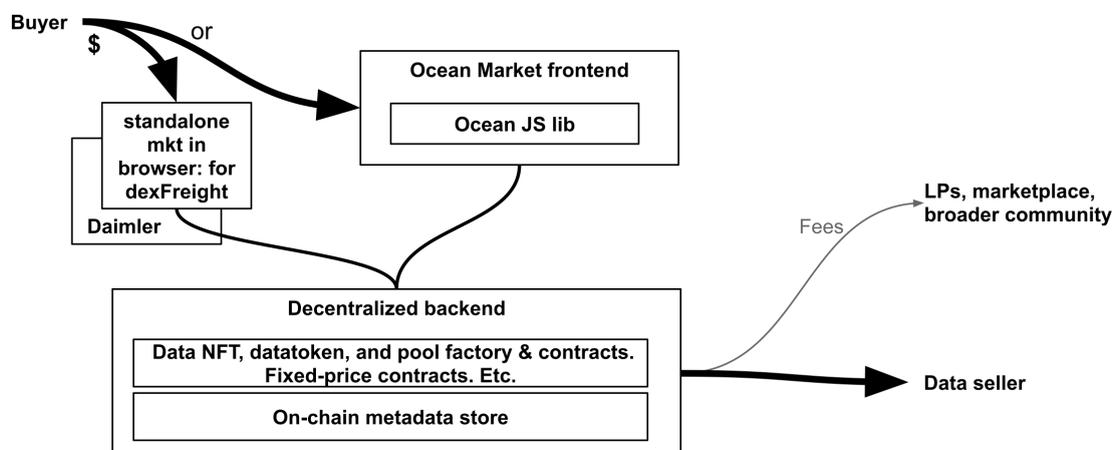


Figure 4: Ocean Marketplaces share a decentralized backend

### 3.6.3. Ocean Marketplace Actions

**Functionality covered.** A marketplace's core functionality is about connecting buyers to sellers for given assets: to make the assets discoverable, and buying/selling an asset of interest. For smoother user flow, Ocean Market supports adjacent functionality: publishing the asset in the first place, and consuming it. Each subsection will cover these, in the order that it would happen, with a focus on Ocean Market.

**Action: publish dataset.** When the user (publisher) clicks on "Publish", they end up here. They start to fill out metadata, including Title and Description. The publisher then provides the url of where the data asset can be found. This url gets stored encrypted and on-chain. When a buyer later consumes a datatoken, that url will be decrypted. The publisher then fills out price information. It may be fixed price or dynamic (automatic). If automatic, they add liquidity as desired to be in line with their target price. Finally, they hit "publish" and Ocean Market will invoke blockchain transactions to deploy a datatoken contract, and publish metadata on-chain.

**Action: discovery.** Ocean Market will have thousands of data assets. To help discovery, Ocean includes support for browsing, searching, and filtering data assets.

**Action: buy/sell dataset.** Here, a buyer comes to Ocean Market and connects their wallet. Their wallet has some OCEAN. The buyer clicks the "buy" button; then Metamask pops up and asks for the buyer to confirm a transaction to swap OCEAN tokens for 1.0 datatokens. The buyer confirms, and the swap happens on-chain. Now, the buyer now has 1.0 more datatokens in their wallet.

**Action: Consume Dataset.** Here, a datatoken owner comes to Ocean Market and connects their wallet. They go to the appropriate sub-page with the asset that they own. They consume the dataset by clicking the "use" button. They follow the prompts to end up with a downloaded dataset, or to get results of bringing compute to data.

### 3.6.4. Developer Tools & 3rd Party Marketplaces

Ocean Market is just one data marketplace. We envision *many* data marketplaces. We can catalyze this, by making it easy for developers to create their own marketplaces. Using Ocean JavaScript or Python libraries, each of the following is 1–3 lines of JS or PY code:

- Create a data asset (provision data service, deploy datatoken contract, add metadata, mint datatokens)
- Create a fixed-price market
- Buy datatokens
- Submit a datatoken and consume a data asset

The libraries interface to Ocean smart contracts. Ocean Provider is a support tool to provision data assets, and Ocean Aquarius and to help store data on-chain and to query metadata (with the help of a local cache).

With these tools in place, there are two main ways that a developer can build an Ocean-based data marketplace: (1) Fork Ocean Market, which uses Ocean tools, and (2) Build up their own marketplace using Ocean tools more directly (eg Javascript library).

### 3.6.5. Sustainability & Fees

A primary aim is to help **marketplace operators** make money. A data market can readily support many actions: publish, pool-based swap, fixed-rate exchange, consuming (market-side or Provider side). Whenever one of these actions is performed on the market, the market earns a cut. Therefore the market is incentivized to have lots of these actions performed on it, via traditional startup traction activities (improve UX; improve asset supply; etc).

In addition: **data publishers** can make money by publishing useful datasets that people buy & consume. They can also sell their base IP (data NFT). **Liquidity providers** can make money by taking a cut of swaps in datatoken pools. Finally, the Ocean Community takes a small cut to help long-term sustainability of the Ocean ecosystem.

Fee details are at: <https://docs.oceanprotocol.com/core-concepts/fees>

### 3.6.6. IP Violations and Sensitive Data

Ocean tooling gives a path for Ocean-powered data markets to handle IP<sup>5</sup> violations and sensitive data.

**Here are some challenges.** What happens when someone publishes a dataset that isn't theirs? What happens when lawyers send a DMCA takedown to your market? How do laws of a jurisdiction (e.g. USA, Germany) relate to blockchain? What happens if someone publishes PII<sup>6</sup>? These are all thorny questions.

Towards a solution, here's a key insight. It's the dapp - e.g. Ocean Market or a third-party market - that is run *in a given jurisdiction*, typically with a centralized server running the backend (metadata cache, serving up webpage). This has practical implications: the party ensures that their dapp follows the laws of their jurisdiction.

**Here's the specific solution.** The "Ocean Market Purgatory Process" is a process to handle IP violations and sensitive data [\[OceanPurgatory2022\]](#) in Ocean Market. It's for Ocean Market, which is run by Ocean Protocol Foundation (OPF); 3rd party data markets can leverage this process, or use whatever other process they choose.

Here's an overview of the process. Purgatory can be for a data asset or an actor, as shown in Figure 5 and Figure 6 respectively.

- **Purgatory for a data asset** is the state in which a published data asset is tagged as "in purgatory" in Ocean Market. Being in purgatory has implications into how the asset is displayed in Ocean Market, and what actions are permitted to be performed on the asset. Once in purgatory, the asset (actor) may stay there or leave purgatory if certain conditions are fulfilled.

---

<sup>5</sup> IP = Intellectual Property. This includes copyright and trademark IP.

<sup>6</sup> PII = Personally Identifiable Information. This is enough info to be able to identify a unique person. GDPR European data protection laws (GDPR) mandate the "right to be forgotten" for PII.

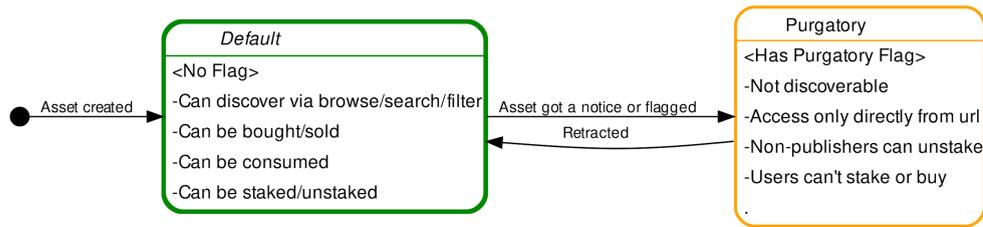


Figure 5: Purgatory for a data asset

- Purgatory for an actor** is a state in which an actor (Ethereum address) is tagged as “in purgatory” in Ocean Market. Being in purgatory has implications into how the actor profile is displayed in Ocean Market, and what actions the actor is permitted to be perform in Ocean Market. Once in purgatory, the actor may stay there, or leave purgatory if certain conditions are fulfilled.

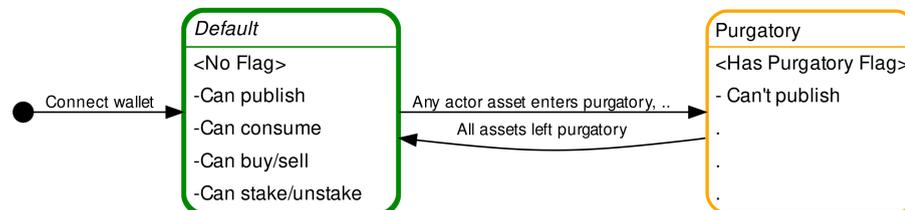


Figure 6: Purgatory for an actor

**A data asset may enter purgatory** if it receives a formal request for takedown due to sensitive data, IP violation, or otherwise. A data asset can either stay in purgatory indefinitely or revert back to default, according to specific conditions.

**An actor enters purgatory** if they have published an asset that’s been put into purgatory, or if they rug pull. The actor can leave purgatory if assets by the actor are on longer in Purgatory; the actor submits a request in writing to OPF, who then decides at its discretion whether to revoke purgatory status.

This process is modeled on GitHub’s approach. The provenance of assets and actors entering purgatory is all on git, stored in a public repo on Github. This makes it precise and transparent. The file `list-assets.json` lists all assets in Purgatory. It follows the following schema.

```
{
  "did": "did:op:0000",
  "reason": "stolen data"
}, ...
```

There’s a similar schema for accounts (actors) in Purgatory. More information is at [\[OceanPurgatory2022\]](#).

### 3.6.7. Benefits of Ocean Data Marketplaces

Ocean marketplace tools make it easy to build & launch data marketplaces.

Ocean-based data marketplaces have these characteristics:

- Interoperability** - data assets being bought and sold are ERC20 tokens on the Ethereum mainnet and other EVM-compatible chains, which play well with the broader Ethereum ecosystem.

- *Don't need login* - users just connect their Web3 wallet (Metamask, etc). Therefore to buy or sell datatokens, they're in and out in 2 minutes. This big UX improvement feels similar to DEXes, versus traditional CEXes<sup>7</sup>.
- *Non-custodial & decentralized* - no centralized middlemen controls the data NFTs & datatokens. No single point of failure.
- *Censorship-resistant, with flexibility* - by default, everyone can transact with the marketplace on the same terms, regardless of their personal identifying characteristics. Or, to meet data regulations or KYC, there is the option of fine-grained permissions.
- *Buy & sell private data while preserving privacy* - using Ocean Compute-to-Data. Data won't leave the premises. This also gives sellers the option to make data *exclusive (in an economic sense)* which can give data a pricing premium.
- *Provenance* - sellers and buyers benefit from the auditability of purchase transactions (using e.g. Etherscan).
- *Monetization* - marketplace has the option to take a commission on sales. This helps to ensure that data marketplace businesses can be built that can sustain themselves and grow over time.

The essay [\[McConaghy2020f\]](#) elaborates on more aspects of Ocean-powered marketplaces, especially Ocean Market. Also, [\[McConaghy2020j\]](#) elaborates on selling data, respectively.

### 3.7. Ocean Tools: Compute-to-Data

Ocean Compute-to-Data provides a means to share or monetize one's data while preserving privacy. This section expands on this.

#### 3.7.1. Motivation

Private data is data that people or organizations keep to themselves, or at least *want to* keep to themselves. It can mean any personal, personally identifiable, medical, lifestyle, financial, sensitive or regulated information.

Privacy tools are about asymmetric information sharing: get info to the people you want, for their benefit, while ensuring that others don't see the info.

**Benefits of Private Data.** Private data can help research, leading to life-altering innovations in science and technology. For example, more data improves the predictive accuracy of modern AI models. Private data is often considered the most valuable data because it's so hard to get, and using it can lead to potentially big payoffs. It's often considered as a competitive, or even decisive, advantage in their market by companies.

**Risks of Private Data.** Sharing or selling private data comes with risk. What if you don't get hired because of your private medical history? What if you are persecuted for private lifestyle choices? Large organizations that have massive datasets know their data is valuable — and potentially monetizable — but do not pursue the opportunity for risk of data escaping and the related liability.

**Resolving the Tradeoff.** There appears to be a tradeoff between benefits of using private data, and risks of exposing it. What if there was a way to get the benefits, while minimizing the risks? This is the idea behind Compute-to-Data: let the data stay on-premise, yet allow 3rd parties to run specific compute jobs on it to get useful analytics results like averaging or building an AI model. The analytics results help in science, technology, or business contexts; yet the compute is sufficiently "aggregating" or "anonymizing" that the privacy risk is minimized.

---

<sup>7</sup> CEX = Centralized Exchange

### 3.7.2. Conceptual Working

Figure 7 illustrates Compute-to-Data conceptually. Alice the data scientist goes to a **data marketplace** and purchases access to private data from seller Bob. She runs her AI **modeling algorithm** (which Bob has approved) on Bob's **private data** to **privately train a model**, which Bob also stores privately. She then runs the trained **private model** on new input data to get **model predictions**. Those predictions are the only data she sees. Everyone is satisfied: Alice gets predictions she wants, and Bob keeps his data private.

At the heart, there are **datatokens** which grant access to run compute next to the data. These datatokens can be used within a marketplace context (like described) or other contexts.

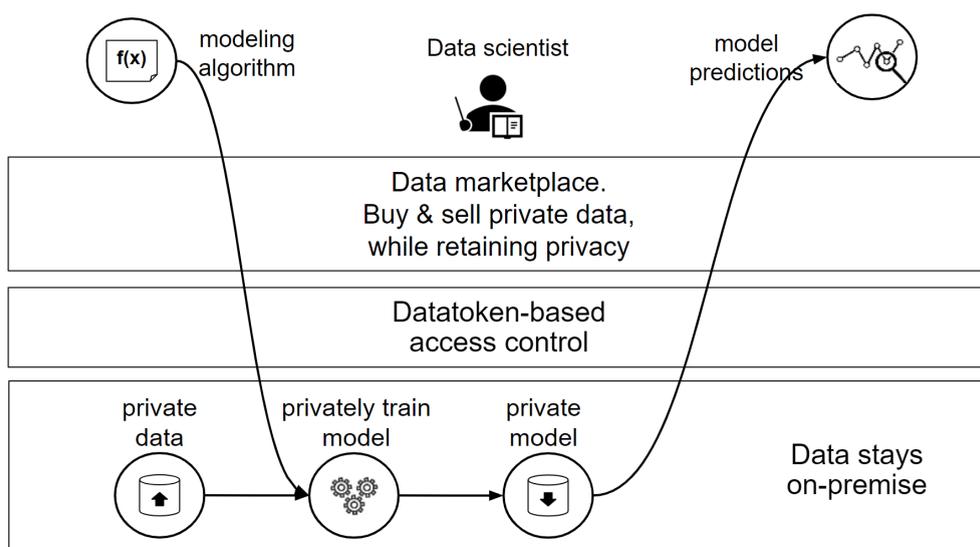


Figure 7: Ocean Compute-to-Data

### 3.7.3. Compute-to-Data Flow Variants

This section describes variants of how Compute-to-Data may be used. For further detail yet, we refer the reader to [\[McConaghy2020c\]](#).

In these variants, there is still model training next to the data:

- Alice is able to download the trained model, i.e. it can leave Bob's premises. This will happen if Bob believes the model has low risk of leaking personally identifiable information (PII), such as being a linear model or a small neural network.
- Alice learns a model across *many* data silos. This is "Federated Learning."

In the following, the compute that is run next to the data is *not* for training a model, but something else.

- A simpler "aggregating" function is run next to the data, such as an average, median, or simple 1-d density estimation. This means that Compute-to-Data is useful for simpler business intelligence (BI) use cases, in addition to more complex artificial intelligence (AI) use cases.
- An aggregating function is computed across *many* data silos. This is "Federated Analytics."
- A synthetic-data generation algorithm is run next to Bob's data. Alice downloads the synthetic data, which she then visualizes or trains models.

- A *hash* is computed for each {input variable, input value} combination of Bob’s data, where the compute is done next to the data. Hashing naturally anonymizes the data. Alice downloads this hashed data and trains the model client-side. This is called “Decoupled Hashing.”
- Random noise is added to Bob’s dataset, sufficiently so for the data to be considered anonymized. Alice downloads this partly-randomized data and trains the model client-side. This is a variant of “Differential Privacy.”

#### 3.7.4. Share, or Monetize

Compute-to-Data is meant to be useful for data sharing in science or technology contexts. It’s also meant to be useful for monetizing private data, while preserving privacy. This might look like a paradox at first glance but it’s not! The private data isn’t directly sold; rather, specific access to it is sold, access “for compute eyes only” rather than human eyes. So Compute-to-Data in data marketplaces is an opportunity for companies to monetize their data assets.

#### 3.7.5. Compute-to-Data Architecture

**New actors.** Ocean Protocol has these actors: **Data Providers**, who want to sell their data; **Data Consumers**, who want to buy data; and **Marketplaces**, to facilitate data exchange. Compute-to-Data adds a new actor, the **Compute Provider**. The Compute Provider sells compute on data, instead of data itself.

**New Components.** Ocean technology has several components. Operator Service and Operator Engine were introduced for Compute-to-Data.

- **Operator Service** — a microservice in charge of managing the workflow and executing requests. It directly communicates and takes orders from Provider (the data provider’s proxy server) and performs computation on data, provided by Provider.
- **Operator Engine** — a backend service in charge of orchestrating the compute infrastructure using Kubernetes as a backend. Typically, the Operator Engine retrieves the workflows created by the Operator Service in Kubernetes. It also manages the infrastructure necessary to complete the execution of the compute workflows.

**New Asset Type.** Before, datasets were the only asset type in metadata (DDO). Compute-to-Data introduces a new asset type — algorithm, which is a script that can be executed on datasets.

For further detail, [\[Patel2020\]](#) provides a worked example and further references on Compute-to-Data.

#### 3.7.6. Marketplaces and Compute-to-Data

Marketplaces can allow their users to publish datasets with Compute-to-Data enabled. Some marketplaces may even require it.

Marketplaces choose what exact compute resources they want to make available to their end users within a K8s cluster, even having them choose from a selection of different images and resources.

Likewise, marketplaces can choose and restrict the kind of algorithm they want to allow their users to run on the datasets in a marketplace.

### 3.7.7. Trusting Algorithms

Only trust in a narrow facet is required: does the algorithm have negligible leakage of PII? For example, a simple averaging function aggregates data sufficiently to avoid leaking PII. AI algorithms also aggregate information.

The data owner typically chooses which algorithms to trust. It's their judgement call. They might inspect the code and perhaps run it in a sandbox to see what other dependencies it causes, communications it invokes, and resources it uses. Therefore it's the same entity that risks private data getting exposed and chooses what algorithm to trust. It is their choice to make, based on their preference of risk vs. reward.

This also points to an opportunity for marketplaces of vetted algorithms: Ocean marketplaces themselves could be used, where liquidity provided is a proxy for quality and trust of the algorithm. Like with all pools, anyone can provide liquidity. This may be quite powerful, as it creates a "data science" side of the market.

### 3.7.8. Benefits of Compute-to-Data

Compute-to-Data has privacy benefits and other benefits as well:

- **Privacy.** Avoid data escapes, never leak personal or sensitive information.
- **Control.** Data owners retain control of their data, since the data never leaves the premises.
- **Huge datasets.** Data owners can share or sell data without having to move the data, which is ideal for very large datasets that are slow or expensive to move.
- **Compliance.** Having only one copy of the data and not moving it makes it easier to comply with data protection regulations like GDPR [\[GDPR2020\]](#).
- **Auditability.** Compute-to-Data gives proof that algorithms were properly executed, so that AI practitioners can be confident in the results.

### 3.7.9. Relation to other Privacy-Preserving Technologies

Compute-to-Data is complementary to other technologies such as encryption/decryption, Multi-Party Compute, Trusted Execution Environments, and Homomorphic Encryption. This compatibility and Ocean's "tools" framing helps make it easy to adopt. [\[McConaghy2020c\]](#) has details.

## 3.8. Ocean Tools: Fine-Grained Permissions

### 3.8.1. Introduction

Ocean Fine-Grained Permissions [\[Hewitt2021\]](#) enable publishers and data marketplace owners to restrict access to publishing and consuming in an enterprise-oriented data marketplace.

A large part of Ocean is about **access control**, of the decentralized variety. Ocean's baseline approach is *datatokens*: a consumer can *access* a resource (e.g. a file) by redeeming 1.0 of the datatokens for that resource.

Enterprises and other users often need more precise ways to specify and manage access. Consider a medical data use case, where only an appropriately credentialed EU researcher can **browse** available medical datasets, or **download** the datasets themselves. Or what if one wanted to restrict to employees of a specific country, or consortium, region, or a combination thereof? Questions like this were often asked by our enterprise collaborators.

This means Ocean must address how to control access more precisely, at these two levels:

1. **Market-level fine-grained permissions** for *browsing, publishing*, etc within a marketplace frontend
2. **Asset-level fine-grained permissions** on *consuming* the asset itself

Ocean supports both.

The next two sections describe the high-level approach for each. For further detail yet, we refer the reader to the original article [\[Hewitt2021\]](#) and related Ocean documentation [\[OceanDocsPermissions2022\]](#).

### 3.8.2. Market-Level Permissions

Back to the medical data use case: only an appropriately credentialed EU researcher can use the marketplace data. How to specify who can access it?

A first idea would be to have a list of all the “allowed” people. But this has problems. Imagine an enterprise publishing 1000 data assets, and for each asset having a list of all its employees. If someone new is hired, they’d have to update all 1000 assets.

Fortunately, there’s a cleaner way. Access is based on a *role*, specified as a *credential*. E.g. “is an employee of x”. To continue the example above: new hires might get the credential when they join the company. Since they have the credential, they have permission. That’s it :) This credential-based approach is known in the data industry as **Role-Based Access Control (RBAC)** [\[RBAC2022\]](#) .

For market-level permissions, Ocean implements RBAC via an “RBAC server”. It implements restrictions at the user level, based on the user’s role (credentials).

The RBAC server is run & controlled by the marketplace. Therefore permissions at this level are defined by the marketplace.

### 3.8.3. Asset-Level Permissions

Let’s go back to the medical data use case, where only an appropriately credentialed EU researcher can **download** the datasets themselves.

How do we handle this? There’s a tension:

- Datatokens on their own aren’t enough — it would mean that anyone with 1.0 datatokens could access that data.
- Yet we want to retain datatokens approach since they enable Ocean users to leverage existing crypto infrastructure to get data wallets, data exchanges, data unions, and more out-of-the-box. The thousands of man-years of effort dedicated to create these crypto tools can be used directly as tools for a data ecosystem.

We can resolve this tension by drawing on the following mental model:

*Imagine going to an age 18+ rock concert. You can only get in if you show both (a) your concert ticket and (b) an id showing that you’re old enough.*

We can port this model into Ocean, where (a) is a datatoken, and (b) is a *credential*. The datatoken is the baseline access control. It’s fungible, and something that you’ve paid for or had shared to you. It’s independent of your identity. The credential is something that’s a function of your identity or your roles in society (the RBAC part).

Ocean now supports fine-grained permissions for asset-based access control. Here’s how: a consumer can only access the resource if they have 1.0 datatokens and one of the credentials in the “allow” list of the DDO.

Ocean also has complementary deny functionality: if a consumer is on the “deny” list, they will not be allowed to access the resource. The rules combine as follows:

*Access = allowed — denied.*

*Allowed = everyone by default, or allow list if it has entries.*

*Denied = entries in deny list.*

Initially, the only credential supported is Ethereum public addresses. To be fair, it’s more a pointer to an individual, not a credential; but it has a low-complexity implementation so makes a good starting point. For extensibility, the Ocean metadata schema enables the specification of other types of credentials like W3C Verifiable Credentials and more. When this gets implemented, asset-level permissions will be properly RBAC too.

Since asset-level permissions are in the DDO, and the DDO is controlled by the publisher, asset-level restrictions are controlled by the publisher.

## 4. Subblock: OceanDAO

### 4.1. Introduction

The previous section described Ocean Tools to power a data ecosystem. We now discuss the next subblock in the Ocean System: OceanDAO. It addresses the questions:

- “How to kickstart growth?” and
- “How can the system not only self-sustain but actually improve over the decades, without guidance by a centralized actor?”

Our answer to this is OceanDAO: a DAO with its own treasury, active treasury management, and grants program. OceanDAO closes the loop with the rest of the Ocean System, such that growth in Ocean usage leads to more funding to teams, to work on further growing usage. This helps both short-term growth and long-term sustainability.

### 4.2. OceanDAO Components

OceanDAO has an initial treasury from the 51% of unminted OCEAN.

It converts some to ETH, DAI, etc by selling time-locked OCEAN (bonds).

It grows its treasury via income from Network Revenue and from active treasury management (LPing into OCEAN-ETH and OCEAN-DAI, yield farming, etc).

It spends its treasury towards rewards (e.g. Data Farming) and grants.

### 4.3. OceanDAO Implementation

Many tools for DAOs are emerging. At the smart contract level, these include Aragon [\[Aragon2020a\]](#), Moloch V2 [\[Turley2020\]](#), DAOStack [\[DaoStack2020\]](#), Colony [\[Colony2020\]](#), and Collab.Land [\[CollabLand2020\]](#). Higher-level interfaces/apps for soft-signaling or voting include Snapshot, Boardroom, Pokemol, Daohaus, Discourse, Discord, and Telegram.

The DAO & governance space is evolving rapidly. Ocean will need to have flexibility to be able to learn over time. Therefore rather than deploying one “perfect” DAO a single time and hope that it stays perfect, we plan to “bake slowly” to give ample opportunity to change things.

OceanDAO was deployed live on Nov 30, 2020 [\[Napheys2020\]](#), initially as just a grants DAO.

Proposals are posted on a Ocean Port (a Discourse Forum), discussion has text chat on Ocean Discord

and videochat in weekly Town Halls, and voting on Snapshot. The OceanDAO wiki is a unifying connector of these tools [\[OceanDAOwiki\]](#).

OceanDAO's scope is expanding over time. During 2022 this includes expansion to holding its own treasury, active treasury management, and implementation of Rewards programs.

## 5. Ocean Data Farming

### 5.1. Overview

Ocean **Data Farming** is a program to incentivize a supply of **relevant and high-quality data assets**.

ODF aims to maximize the **data supply reward function (RF)**, which is a function of liquidity added, dataset usage, and potentially more. In other words, liquidity providers (LPs) can earn extra OCEAN rewards for providing liquidity, and amount earned will be multiplied by usage of the datasets they stake on, and more.

We will start with an initial RF and a program budget of OCEAN over a time interval (e.g. 3 months). Each week at a pre-set time, the program managers will:

- Off-chain, calculate contribution of each actor to the RF for that week (actor = Ethereum address).
- Manually airdrop OCEAN to each actor proportional to their contribution, the total being that week's budgeted OCEAN.
- Declare the RF for use in the following week, to fine-tune towards the most relevant and high-quality datasets, and away from "gaming" that is against the spirit of the program.

During the week, the program managers will:

- Gather feedback from the community on possible ways to improve the RF, using soft signaling<sup>8</sup> in Discord or similar. Later generations may use OCEAN token voting.

The OCEAN will budget may draw from OPF treasury and (if the community votes for it) OceanDAO funds.

This design allows the program managers to start simple, then to quickly learn and improve the RF every week. This design draws inspiration from the successful Balancer Liquidity Mining Program [\[Martinelli2020\]](#). If the initial ODF run (over the time interval) is successful, it may be repeated.

### 5.2. Reward Function

This section describes the RF.

$RF_{ij}$  is the (non-normalized) contribution for actor  $i$  on datatoken  $j$ :

$$RF_{ij} = S_{ij} * D_j$$

where:

- $S_{ij}$  = actor  $i$ 's OCEAN stake in data asset  $j$  = (actor's # BPTs<sup>9</sup> in datatoken  $j$ 's pool / total # BPTs in pool) \* (# OCEAN in pool)<sup>10</sup>

---

<sup>8</sup> E.g. giving a thumbs-up or thumbs-down to questions posted in Discord.

<sup>9</sup> BPT = Balancer Pool Token. People receive BPTs proportional to the OCEAN or datatoken liquidity they add to a pool. Conversely, when they remove liquidity, they are essentially selling BPTs for OCEAN and datatokens.

<sup>10</sup> In the first iteration, rewards will only go to liquidity in Ocean-tweaked Balancer pools. Over time, we anticipate this to expand to other exchanges.

- $D_j$  = # times data asset  $j$  has been consumed in the last week (= # consume transfers to the datatoken  $j$ 's Provider)

The first term is  $S_{ij}$ . It reflects the actor's belief in the relevance, quality, or potential usage of the datatoken. This incentivizes Providers to publish relevant data assets. The second term,  $D_j$ , is a direct measure for the usage of the data asset, and a proxy for its relevance or quality. This  $RF_{ij}$  can be summarized as a binding of *predicted* popularity \* *actual* popularity in terms of their orders-of-magnitude. As discussed, this is the *initial* RF. We anticipate feedback from the community to further improve it.

To help long-term thinking and drive value to the protocol, we are considering lockup periods to ensure that rewards are not immediately sold by traders [Jain2020]. It could even be "longer lockup, longer reward" [Fiskantes2020], like in Filecoin's ICO [Coinlist2017].

Astute readers may recognize this design from the 2018 and 2019 versions of the Ocean whitepaper, to maximize the supply of data. That's correct. There are two main differences. We've changed it from smart contracts in the core of the system, into a program to catalyze growth run manually and off-chain for rapid learning and iterations. Also, the rewards formula has a simpler implementation, due to using staking inside datatoken pools and datatoken transfer calls.

The blog post "Ocean Data Farming is Launching" [McConaghy2022a] has further details.

## 6. On OCEAN Token Model

### 6.1. Introduction

This section summarizes the OCEAN token from a crypto-economic perspective.

### 6.2. Token Value $\propto$ Volume

To ensure long-term funding for the Ocean community, **we want OCEAN health to increase as usage volume increases. Ocean does this; here's how.**

- Here's the loop: more usage of Ocean tools in the Data Ecosystem leads to more OCEAN being staked, leading to more OCEAN demand, growing \$OCEAN.
- More usage also leads to more Network Revenue, which goes to (i) burning and (ii) OceanDAO. Burning OCEAN reduces supply, to grow \$OCEAN. Funds go through OceanDAO to workers who have the mandate to grow usage of Ocean tools. And the loop repeats.

### 6.3. Work Token (and Governance Token)

"Work" is when actors perform services for the token ecosystem that add value to the system<sup>11</sup>.

Ocean incentivizes for work to happen in OceanDAO governance, Ocean Workers block, and Ocean marketplace tools. This gives OCEAN some **work token** dynamics [Samani2018].

**OceanDAO Governance.** In Ocean, governance is about directing the flow of resources towards the teams and projects that offer the best chance for growth. OCEAN holders are incentivized to do work to learn more about each team and project proposal, and then use OceanDAO to vote for the most promising teams/projects (subject to Ocean mission & values). If there aren't enough value-adding projects for the OCEAN funds available, remaining OCEAN gets burned; this therefore sets a baseline

---

<sup>11</sup> This is more general than the idea that one must buy a work token to have access to revenue streams in the network (like a NYC taxi medallion).

for the value-add needed. Note that governance is *not* about specific decisions for protocols or software, it's at a higher level and more directly about value creation.

**Ocean Workers.** Paid by OceanDAO, these teams do a variety of work, including: improving core software (datatokens, etc), building and improving applications, outreach and community building, and unlocking data for use in Ocean. All of these activities drive value to the Ocean ecosystem (on average) because the work has been filtered by OceanDAO towards the value-adding activities.

**Ocean marketplace tools.** There are several ways for the Ocean ecosystem to gain value via “work” inside the “data ecosystem” box of Figure 1.

- **More liquidity** in datatoken pools gives lower slippage, for a better data buyer experience, which in turn drives volume, and in turn Ocean value. The promise of **LP transaction fees** incentivizes people to add OCEAN liquidity to OCEAN-datatoken pools.
- **Curation of assets** (=amount staked in the corresponding datatoken pool) allows data buyers to surface the best assets more easily, which in turn drives volume, and in turn Ocean value. To have the best returns, LPs are incentivized to **seek datatoken pools with the highest volumes**, which tend to be the ones with the best datasets. That is, they're curating.
- **Referrals** bring new users to Ocean tools, which in turn drives volume, and in turn Ocean value. Because LPs / stakers / curators get fees on transactions, they are incentivized to refer people to the pools that they have staked in.
- **More marketplaces** with good volumes drives total Ocean volume, and in turn Ocean value. The promise of **marketplace fees** incentivizes people to run their own marketplaces and grow their volumes.

**Ocean Data Farming.** This program incentivizes workers to add data assets, use data assets, and add liquidity to data assets. Future versions may also use OCEAN-based voting to govern updates to their reward formulae.

#### 6.4. Burning

First, 5% of all Network Revenue is burned. This ensures that OCEAN holders will always see some deflationary actions, and as Network Revenue increases, deflation increases. The 5% value was chosen based on running various possible values in agent-based simulation [[TokenSPICE2020](#)].

Second, for a given funding period, OceanDAO burns all non-allocated funds.

#### 6.5. veOCEAN

veOCEAN is a means to lock OCEAN to be able to vote in OceanDAO and earn ecosystem rewards. veOCEAN holders can earn passively, but can earn more by actively allocating veOCEAN to data assets that have high data consume volume (DCV).

The blog post "Introducing veOCEAN" [[McConaghy2022b](#)] has further details.

#### 6.6. User Acquisition

Special OceanDAO projects like Data Supply Mining are not core to the Ocean token design itself, but spend some OCEAN resources in order to help acquisition of users, data, liquidity, and the like.

#### 6.7. On OCEAN as Unit-of-Exchange

OCEAN can be used as a unit-of-exchange to buy and sell datatokens, but it's not mandated. Being ERC20 tokens, datatokens may be transferred in return for whatever payment is desired, such as fiat, DAI, or ETH. Community participants may add capabilities like this as they see fit.

OCEAN is the default unit-of-exchange in Ocean Market and in Ocean libraries, for both fixed-price and auto-pricing scenarios (the latter also defaults to staking OCEAN). While Ocean Market transacts in OCEAN, it also displays the price in USD and EUR. Community participants may build data marketplaces that transact in fiat or other crypto-tokens.

The essay [\[McConaghy2020g\]](#) elaborates on the Ocean token model further.

## 7. Ocean Applications

### 7.1. Decentralized Orchestration

Here's an example compute pipeline: input raw training data → clean the data → store cleaned data → build model → store model → input raw test data → run predictions → store predicted result  $y_{test}$ . Leveraging Ocean, a developer can write Solidity code [\[Solidity2020\]](#) to define a pipeline and execute it, i.e. do “decentralized orchestration”:

1. Write a smart contract that uses various data services. There's a data NFT & datatoken for access control of each data service.
2. Deploy the smart contract to the EVM chain.
3. For each different datatoken, do a tx that approves the required amount of datatokens to the SEA smart contract.
4. Finally, do a transaction that makes the actual call to the smart contract.

With a Set Protocol, ERC998 or similar to have a “basket” datatoken that holds all the necessary sub-datatokens in the compute pipeline, step 3 becomes simpler yet.

A major sub-application is decentralized federated learning (decentralized FL). In traditional FL, a centralized entity (e.g. Google) must perform the orchestration of compute jobs across silos. So, PII can leak to this entity. OpenMined [\[OpenMined2020\]](#) could decentralize orchestration, using Ocean to manage computation at each silo in a more secure fashion [\[McConaghy2020c\]](#).

### 7.2. Data Wallets: Data Custody & Data Management

Data custody is the act of holding access to the data, which from an Ocean framing is simply holding data NFTs or datatokens in wallets. Data management also includes sharing access to data, which in Ocean is simply transferring data NFTs or datatokens to others.

With data NFTs & datatokens following the ERC721 & ERC20 standards, we can leverage existing wallets for ERC721 & ERC20. This includes browser wallets (e.g. Metamask), mobile wallets (e.g. Argent, Pillar), hardware wallets (e.g. Trezor, Ledger), multi-sig wallets (e.g. Gnosis Safe), institution-grade wallets (e.g. Riddle & Code), custodial wallets (e.g. Coinbase Custody), and more.

Wallets may get tuned specifically for datatokens as well, e.g. to visualize datasets, or long-tail token management (e.g. holding 10,000 different datatoken assets).

Existing software could be extended to include data wallets. For example, Brave browser has a built-in crypto wallet that could hold datatokens. There could be browser forks focused on datatokens, with direct connection to user browsing data. Integrated Development Environments (IDEs) for AI like Azure ML Studio [\[Azure2020\]](#) could have built-in wallets to hold & transfer datatokens for training data, models as data, and more. Non-graphical AI tools could integrate; such as scikit-learn or TensorFlow Python libraries using a Web3 wallet (mediated with Ocean's Python library).

As token custody continues to improve, data custody inherits these improvements.

### 7.3. Data Auditability

Data auditability and provenance is another goal in data management. Thanks to data NFTs & datatokens, blockchain explorers like Etherscan [\[Etherscan2020\]](#) now become data audit trail explorers.

Just as CoinGecko [\[CoinGecko2020\]](#) or CoinMarketCap [\[CoinMarketCap2020\]](#) provide services to discover new tokens and track key data like price or exchanges, we anticipate similar services to emerge for data NFTs & datatokens. CoinGecko and CoinMarketCap may even do this themselves, just as they've done for DeFi tokens.

### 7.4. Data DAOs: Data Co-ops and More

Decentralized Autonomous Organizations (DAOs) [\[DAO2020\]](#) help people coordinate to manage resources. They can be seen as multi-sig wallets, but with significantly more people, and with more flexibility. DAO technology is maturing well, as we reference in the OceanDAO section. A “data DAO” would own or manage data NFTs or datatokens on behalf of its members. The DAO could have governance processes on what data NFTs or datatokens to create, acquire, hold, sell / license, and so on.

Here are some applications of data DAOs:

**Co-ops and Unions (Collective Bargaining).** Starting in the early 1900s, thousands of farmers in rural Canada grouped into the SWP [\[SWP2020\]](#) for clout in negotiating grain prices, marketing grain, and distributing it. Labour unions have done the same for factory workers, teachers, and many other professions. In [\[Ibarra2018\]](#), the authors suggest that data creators are currently getting a raw deal, and the solution is to make a labour union for data. A data DAO could be set up for collective bargaining, as a “data coop” or “data union”. For example, there could be a data coop with thousands of members for location data, using FOAM proof-of-location service [\[McConaghy2020a\]](#).

**Manage a single data asset.** There could be a DAO attached to a single data asset. One way is: create a Telegram channel dedicated to that dataset. You can only enter the Telegram channel if you have 1.0 of the corresponding datatokens (inspired by Karma DAO [\[Lee2020\]](#)). This can also be for Discord, Slack, or otherwise.

**Datatoken pool management.** There could be a data DAO to manage a datatoken pool's weights, transaction fees, and more, leveraging Balancer Configurable Rights Pools [\[Balancer2020b\]](#) (inspired by PieDAO which does this for a pool of DeFi assets [\[Delmonti2020\]](#)).

**Index Funds for Data Investments.** Using e.g. Melon [\[Melon2020\]](#), an investment product can be constructed to allow people to buy a basket of data assets with the current plethora of mutual and index funds as a guide.

### 7.5. Fractional Ownership

Fractional ownership is an exciting sub-niche of Web3, at the intersection of NFTs and DeFi. IT allows co-ownership of data IP.

Ocean provides two approaches to fractional ownership:

1. **Sharded holding of ERC20 datatokens**, where each ERC20 holder has the usual datatoken rights as described above, e.g. 1.0 datatokens to consume an asset. This comes out-of-the-box with Ocean.

2. **Sharding ERC721 Data NFT**, where each co-holder has right to some earnings against base IP, and co-controls the Data NFT. For example, there's a DAO with the sole purpose to hold the Data NFT; this DAO has its own ERC20 token; DAO members vote with tokens to update Data NFT roles or deploy ERC20 datatokens against the ERC721.

Note: For (2), one might consider doing sharding with something like Niftex. But then there are questions: what rights do the shard-holders get exactly? It could be zero; for example AMZN shareholders don't have the right to walk the hallways of Amazon Inc. just because they hold shares. Secondly, how do the shard-holders control the Data NFT? These questions get resolved by using a tokenized DAO, as described above.

## 7.6. Group-Restricted Data Access

In this operational model, "membership rules" apply for a group. These rules are governed by a verifiable credential, a custom list of allowed Ethereum addresses, otherwise. Ocean implemented these via the "Fine-Grained Permissions" feature described earlier.

These membership rules enable the following applications for data sharing.

**Contests, Hackathons, Impromptu Collaborations.** A group of hackers or data scientists self-organize to try to solve an AI problem, such as a Kaggle competition or a hackathon. They want to be able to easily access each others' data and compute services as they progress, especially if they are working remotely from each other.

**Regulatory Sandboxes.** A government wants to give a means for organizations to run in a "monitored" regulatory environment that the government can observe. The organizations are vetted by the government and may have access to specially designated government compute services.

**Enterprise data access.** An enterprise might make some of its data available to only its employees, but want to be able to use Ocean services available in the broader network.

**Sharing autonomous driving data.** Individuals in each membership company of MOBI [\[MOBI2020\]](#) need to access automotive data from any one of the MOBI member companies. It could be time-consuming and error-prone to specify permission for each member company individually. Furthermore, those permissions will fall out of date if MOBI members are added or removed; and updating the permissions one organization at a time could also be time-consuming or error-prone. This involves two levels of permissions: access of member companies into MOBI, and access of individuals in each member company (enterprise).

**Sharing medical data.** Researchers on European soil that wish to directly access German medical data need to demonstrate that they have been accredited by appropriate authorities. This will usually be through their hospital or university. There could be thousands of researchers accessing the data. As with automotive data, it will be time-consuming and error prone to specify and update permissions for each of these thousands of researchers. This may be two levels of permissions (hospital/university into EU authority; individual into hospital/university), or it may be among hospitals and universities in a more networked fashion.

**Sharing financial data (while preserving privacy).** Small and medium-sized credit unions in the U.S. have a challenge: they don't have large enough datasets to justify using AI. Since the credit unions don't compete with each other, they would find great value to build AI models across their collective datasets.

## 7.7. Unlocking Latent Data of Individuals & Enterprises

Specialized apps could get built to tokenize and earn from latent data assets of individuals or organizations.

**Individuals.** For example, an app that you install on your phone, you give permissions to access data (Compute-to-Data for privacy), then it auto-creates a fixed-price market for your data. After that you earn money from the data on your phone. In a mashup with personal tokens [\[Chaturvedi2020\]](#), these would be your sovereign “personal datatokens” and you could launch your data as an “Personal Data Offering” (PDO).

One could even launch multiple PDOs. There would be one datatoken for each data type generated - smartphone data, smartwatch data, browser data, shopping data, and so on. Then, bundle the datatokens into a collection of personal datatokens (composable datatokens). One could even have a personal data marketplace<sup>12</sup>.

One could do the same with a browser plug-in, to sell the latent data on your browser (cookies, bookmarks, browsing history). New service firms could emerge to help enterprises tokenize and earn from their massive internal data troves.

These data assets might be sold one person or entity at a time. Or, they could be brought into a DAO to pool resources for more marketing and distribution muscle.

**Enterprises.** Large enterprises have massive datasets. They know their data is valuable: they spend millions annually to help protect it and insure it due to hacks. But what if rather than data being a *liability*, data was *assets* on enterprises’ books? Ocean offers this possibility, by making it easy to turn the internal data into fungible assets (via datatokens), while preserving privacy and control (via Compute-to-Data). Data not only becomes a new line of revenue, it becomes a *financial asset* that can be borrowed against to fund growth, and more.

## 7.8. Data Marketplaces

Earlier, we described Ocean Market, an out-of-the-box data marketplace that Ocean offers. Here are some ways that forks of Ocean Market could differentiate.

- **Focus on a given vertical.** For example, the Ocean-based Daimler data marketplace focuses on the mobility vertical. Other verticals include health, logistics, and DeFi.
- **Focus on private data.** A marketplace focusing on Ocean compute-to-data, with features for curation of compute algorithms.
- **Different fee structure.** When a purchase happens in Ocean Market, the default 0.1% fee goes fully to LPs. Variants include: marketplace operator gets a cut, referrers get a cut, charge higher %, charge a flat fee.
- **Novel pricing mechanism.** Many price discovery mechanisms are possible [\[Kuhn2019a\]](#) [\[Kuhn2019b\]](#), including royalties (a % of sales), English / Dutch / Channel auctions [\[Azevedo2018\]](#), or income share agreements like Bowie or Dinwiddie bonds [\[McConaghy2019d\]](#). This may include a novel initial distribution mechanism, as the Initial Data Offerings section elaborates.
- **Different payment means.** Ocean Market takes OCEAN. Variants could take in fiat, DAI, ETH, etc. (Ocean Market may support these directly over time as well. PRs are welcome!)
- **Decentralized dispute resolution** using Aragon Court [\[Aragon2020b\]](#) or Kleros [\[Kleros2020\]](#).
- Different Balancer weighting schemes. For example, make the 90/10 weights shift to 50/50 over time (Liquidity Bootstrapping). Or, surge pricing pools, which have higher fees when

---

<sup>12</sup> Thanks to Simon Mezgec for this suggestion.

there is more demand for liquidity. These could be implemented with Configurable Rights Pools. PRs for Ocean Market are welcome!

Beyond forks of Ocean Market, there is an larger variety of possible marketplaces. Here are some Web2 and Web3 variants.

- **AMM DEXes.** This could be a Uniswap or Balancer-like webapp to swap datatokens for DAI, ETH, or OCEAN.
- **Non-data NFT marketplaces.** Create an NFT marketplace focusing on photos, digital art, music, utility tokens, physical objects, or otherwise. Its USP would be the ability to sell well-defined licenses to single editions, and pools for staking & swapping on those tokens. (DeFi pools are much more fluid than typical NFT auctions.)
- **Order-book DEXes.** It could use Ox, Binance DEX, Kyber, etc. It could leverage platform-specific features such as Ox's shared liquidity across marketplaces.
- **Order-book CEXes.** Centralized exchanges like Binance or Coinbase could readily create their own datatoken-based marketplaces, and to kickstart usage could sell datasets that they've generated internally.
- **Marketplaces in AI tools.** This could be an AI-oriented data marketplace app embedded directly in an AI platform or webapp like Azure ML Studio or Anaconda Cloud. It could also be an AI-oriented data marketplace as a Python library call, for usage in any AI flow (since most AI flows are in Python). In fact, this is already live in Ocean's Python library.
- **"Nocode" Data Marketplace builder.** Think Shopify [\[Shopify2020\]](#) for data marketplaces, where people can deploy their own data marketplaces in just a few clicks.

## 7.9. Data as an Asset Class for DeFi

The data economy is already 377B€ for Europe alone [\[EuroComm2019\]](#). Tokenized data assets have great promise to grow the size of DeFi assets under management.

Data can be securitized and used as collateral [\[McConaghy2019d\]](#). An example is Bowie Bonds, where a fraction of David Bowie's IP (intellectual property) licensing revenue was paid to bondholders. Data is IP. To use it as a financial asset, one must price it. In Bowie's case, the value was established from previous years' licensing revenue. Alternatively, we can establish price by selling data assets in data marketplaces.

As such, data is an asset class. With datatokens, we can onboard more more data assets into each major DeFi service types:

- Data assets can be used as collateral in **stablecoins** and **loans**, therefore growing total collateral. New Order has released **H2O**, a stable asset backed by OCEAN in its V1 (early 2022), and by datatokens themselves in its V2 (late 2022) [\[NewOrder2021\]](#).
- Data assets bought and sold in **DEXes** and **CEXes** contributes to their \$ volume and assets under management (AUM).
- There can be insurance on data assets. As described above, there can be data DAOs, data baskets, and more.

Here comes DataFi.

## 7.10. Data to Optimize DeFi Returns

We can close the loop with data helping DeFi, and vice versa. Specifically: **data can improve decision-making in DeFi to optimize returns**. This will catalyze the growth of DeFi further. Here are some examples:

- **Yield farming.** Data can improve the automated strategies to maximize APR in yield farming. Think yearn.finance robots, but optimized further.
- **Insurance.** Data to lower the risk models in insurance.
- **Loans.** Better prediction of default for under-collateralized loans.
- **Arb bots.** More data for higher-return arbitration bots.
- **Stablecoins.** Assessment of assets for inclusion in stablecoins.

**Data-powered loops.** DeFi [looping](#) techniques further boost returns. For each of the examples above, we envision *loops* of buying more data, to get better returns, to buy more data, and so on. To go even further, we could apply this to data *assets* themselves.

## 7.11. Data Management Platforms for Smart Cities and More

In 2001, the government of Estonia rolled out a data management platform called X-Road [\[XRoad2020\]](#). It then deployed an identity system on top; each citizen received an identity card with a digital signature. Since then, Estonia has rolled out apps for elections, health, taxes, parking, lawmaking, E-Residency and two dozen more government apps, plus 3rd-party apps like banking [\[eEstonia2020\]](#).

Both Ocean and X-Road can be used as digital infrastructure for smart cities' data sharing. X-Road can be seen as a smart city example, since the majority of Estonians live in Tallinn. X-Road has a longer history, but has more centralized control and requires dev-ops effort. While Ocean is younger, by using a global permissionless infrastructure, it has lighter dev-ops requirements. We envision a future with both X-Road and Ocean-based data sharing in smart cities.

Ocean framed as a data management platform can be used not only for data sharing within a city (across citizens), but also within a province/state (across cities), within a nation (across provinces), within an international initiative (across nations, e.g. GAIA-X [\[GAIA-X2020\]](#)), within a company (across employees), and within a multinational enterprise (across national offices).

## 7.12. Composable Data NFTs & Datatokens

Data NFTs & datatokens can be composed into bundles, sets, or groups using ERC998 [\[Lockyer2018\]](#), Set Protocol [\[Set2020\]](#), Melon Protocol [\[Melon2020\]](#), or others. This helps for the following use cases:

- **Group across time.** Package each 10-min chunk of data from the last 24 hours into a single token.
- **Group across data sources.** Package 100 data streams from 100 unique Internet-of-Things (IoT) devices, as a single token.
- **Data baskets for asset management.** Group together 1000 datasets that each have individual (but small) value, to sell as a single asset to others wanting to hold data assets.
- **Data indexes.** Track the top 100 data assets and make it easy for others to invest in those as a single asset, similar to today's index funds.
- **On-chain annotations to metadata.** Use ERC998 in a bottom-up setting to "attach" tags or other information to the data asset. Uses include: reputation given by a marketplace's users,

quality as computed by a marketplace's algorithms, input training data vs output vs a model, industry verticals, and more.

The essay [\[McConaghy2019d\]](#) elaborates further.

## 8. Conclusion

This paper presented Ocean Protocol. Ocean tools help developers build marketplaces and other apps to privately & securely publish, exchange, and consume data. The tools offer an on-ramp and off-ramp for data assets into crypto ecosystems, using data NFTs & datatokens. Composability gives many application opportunities, including data wallets, data marketplaces, data DAOs, and more. Ocean Market is a live reference community marketplace.

Ocean tools are encapsulated in a broader system designed for long-term growth of an open, permissionless Web3 Data Economy. Ocean Data Farming incentivizes a supply of quality data. A key piece is OceanDAO to fund software development, outreach. OceanDAO is funded by revenue from Ocean apps and from Ocean's network rewards.

## 9. References

- [3Box2020] 3Box Team, "3Box: User Data Cloud", 3Box Homepage, last accessed Aug 19, 2020, <https://3box.io/>
- [Aragon2020a] "Aragon", homepage, last accessed Aug. 20, 2020, <https://aragon.org/>
- [Aragon2020b] "Aragon Court", homepage, last accessed Aug. 20, 2020, <https://aragon.org/court>
- [Azevedo2018] Eduardo M. Azevedo, David M. Pennock, and E. Glen Weyl, "Channel Auctions", Aug 31, 2018, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3241744](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3241744)
- [Azure2020] Azure ML Team, "Microsoft Azure Machine Learning Studio", homepage, last accessed Aug. 19, 2020, <https://studio.azureml.net/>
- [Balancer2019] Balancer Team, "Interest-Bearing Stablecoin Pools Without Impermanent Loss", Balancer blog, Oct. 30, 2019, <https://balancer.finance/2019/10/30/interest-bearing-stablecoin-pools-without-impermanent-loss/>
- [Balancer2020a] "Balancer", homepage, last accessed Aug. 21, 2020, <https://balancer.finance>
- [Balancer2020b] Balancer Team, "Configurable Rights Pool", GitHub Repository, published Aug, 2020, <https://github.com/balancer-labs/configurable-rights-pool>
- [Banko2001] M. Banko and E. Brill, "Scaling to Very Very Large Corpora for Natural Language Disambiguation", Proc. Annual Meeting on Association for Computational Linguistics, July, 2001, <http://www.aclweb.org/anthology/P01-1005>
- [@Bitcoin2020] @Bitcoin, Twitter, Apr. 4, 2020 <https://twitter.com/Bitcoin/status/1246482664376414209>
- [Buterin2013] Vitalik Buterin, "Ethereum: A Next-Generation Smart Contract and Decentralized Application Platform", Dec. 2013, <https://ethereum.org/en/whitepaper/>
- [Buterin2018] Vitalik Buterin, "Explanation of DAICOs", Eth.Research Blog, Jan. 2018, <https://ethresear.ch/t/explanation-of-daicos/465>
- [Chaturvedi2020], Vinamrata Chaturvedi, "The Man Who Tokenized Himself Gives Holders Power Over His Life", June 30, 2020, <https://www.coindesk.com/man-who-sells-himself-now-wants-buyers-to-control-his-life>
- [CollabLand2020] "Collab.Land", Homepage, last accessed Aug. 20, 2020, <http://collab.land/>
- [CoinGecko2020], CoinGecko Team, CoinGecko website, last accessed Aug. 19, 2020 <https://www.coingecko.com>
- [CoinMarketCap2020], CoinMarketCap Team, CoinMarketCap website, last accessed Aug. 19, 2020, <https://www.coinmarketcap.com>
- [Coinlist2017], Coinlist Team, "Filecoin Token Sale Economics", fall 2017, [https://coinlist.co/assets/index/filecoin\\_2017\\_index/Filecoin-Sale-Economics-e3f703f8cd5f644aec7ae3860ce932064ce014dd60de115d67ff1e9047ffa8e.pdf](https://coinlist.co/assets/index/filecoin_2017_index/Filecoin-Sale-Economics-e3f703f8cd5f644aec7ae3860ce932064ce014dd60de115d67ff1e9047ffa8e.pdf)
- [Colony2020] "Colony", Homepage, last accessed Aug. 20, 2020, <https://colony.io/>
- [DAO2020] "Decentralized Autonomous Organization", Wikipedia, last accessed Aug. 19, 2020, [https://en.wikipedia.org/wiki/Decentralized\\_autonomous\\_organization](https://en.wikipedia.org/wiki/Decentralized_autonomous_organization)
- [DaoStack2020] "DaoStack", homepage, last accessed Aug. 20, 2020, <https://daostack.io/>
- [Delmonti2020] Alessio Delmonti, "Introducing PieDAO, the asset allocation DAO", Mar. 3, 2020, <https://medium.com/piedao/introducing-piedao-the-asset-allocation-dao-1af9eec5ee4>

[DID2019] “Decentralized Identifiers(DIDs) v0.11: Data Model and Syntaxes for Decentralized Identifiers”, W3C Community Group, Draft Community Group Report 06 February 2019, last accessed Feb. 28, 2019, <https://w3c-ccg.github.io/did-spec/>

[Economist2017] “The World’s Most Valuable Resource is No Longer Oil, But Data”, *The Economist*, May 6, 2017, <https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource>

[eEstonia2020] “E-Estonia”, homepage, last accessed Aug. 19, 2020, <https://e-estonia.com/>

[EthereumEvm2022] “Ethereum Virtual Machine”, Ethereum Foundation website, last accessed Feb. 23, 2022, <https://ethereum.org/en/developers/docs/evm/>

[Etherscan2020] “Etherscan”, homepage, last accessed Aug. 21, 2020, <https://etherscan.io>

[EuroComm2019] European Commission, “Building a Data Economy - Brochure”, Sep. 2019, <https://ec.europa.eu/digital-single-market/en/news/building-data-economy-brochure>

[ERC721] “ERC-721”, Homepage, last accessed Aug. 19, 2020, <http://erc721.org/>

[ERC1167] Peter Murray, Nate Welch, and Joe Messerman, “EIP-1167: Minimal Proxy Contract”, Ethereum Improvement Proposals, no. 1167, June 2018, <https://eips.ethereum.org/EIPS/eip-1167>

[Fiskantes2020] @Fiskantes, Twitter, Aug 12, 2020, <https://twitter.com/fiskantes/status/1293372786002538496?s=11>

[GAIAx2020] “GAIA-X”, homepage, last accessed Aug. 21, 2020, <https://www.data-infrastructure.eu/>

[GDPR2020] “General Data Protection Regulations”, Wikipedia, last accessed Aug. 21, 2020, [https://en.wikipedia.org/wiki/General\\_Data\\_Protection\\_Regulation](https://en.wikipedia.org/wiki/General_Data_Protection_Regulation)

[Goldin2017] Mike Goldin, “Mike’s Cryptosystems Manifesto”, Google Docs, 2017, <https://docs.google.com/document/d/1TcCeAsBIAoFLWSQWYyhjmTsZCp0XqRhNdGMb6JbASxc/edit>

[Graham2005] Paul Graham, “How to Start a Startup”, Paul Graham Blog, March 2005, <http://www.paulgraham.com/start.html?viewfullsite=1>

[Halevy2009] Alon Halevy, Peter Norvig, and Fernando Pereira, “The Unreasonable Effectiveness of Data”, *IEEE Intelligent Systems* 24(2), March-April 2009, <https://research.google.com/pubs/archive/35179.pdf>

[Hewitt2021] Jamie Hewitt, “Fine-Grained Permissions Now Supported in Ocean Protocol”, Ocean Protocol Blog, Sep. 22, 2021, <https://blog.oceanprotocol.com/fine-grained-permissions-now-supported-in-ocean-protocol-4fe434af24b9>

[Ibarra2018] Imanol Arrieta Ibarra et al, “Should We Treat Data as Labor? Moving Beyond ‘Free’”, *AEA Papers and Proceedings*, 108 : 38–42. DOI: 10.1257/pandp.20181003, <https://www.aeaweb.org/conference/2018/preliminary/paper/2Y7N88na>

[Jain2020] Tushar Jain, “Exploring the Design Space of Liquidity Mining”, Multicoins Capital Blog, Aug 13, 2020, <https://multicoins.capital/2020/08/13/exploring-the-design-space-of-liquidity-mining/>

[KellyCriterion2017] “Kelly Criterion”, Wikipedia, last accessed Feb. 17, 2018, [https://en.wikipedia.org/wiki/Kelly\\_criterion](https://en.wikipedia.org/wiki/Kelly_criterion)

[Kleros2020] “Kleros”, homepage, last accessed Aug. 20, 2020, <https://kleros.io/>

[Kuhn2019a] Erwin Kuhn, “Let’s Talk About Data Pricing — Part I”, Aug. 13, 2019, <https://blog.oceanprotocol.com/lets-talk-about-data-pricing-part-i-bbc9cf781d9f>

[Kuhn2019b] Erwin Kuhn, “Let’s Talk About Data Pricing — Part II”, Aug. 21, 2019, <https://blog.oceanprotocol.com/value-of-data-part-two-pricing-bc6c5127e338>

[Lee2020] Andrew Lee, “Announcing Karma DAO: First-Ever Token-Permissioned Networking Chat Group on Telegram”, July 20, 2020, <https://medium.com/@andrwlee/announcing-karma-dao-first-ever-token-permissioned-networking-chat-group-on-telegram-5feab7a54def>

[Lockyer2018] Matt Lockyer, “ERC-998 Composable Non-Fungible Token Standard”, Apr. 15, 2018, <https://github.com/ethereum/eips/issues/998>

[Martinelli2020] Fernando Martinelli, “Balancer Liquidity Mining Begins”, Balancer Blog, May 29, 2020, <https://medium.com/balancer-protocol/balancer-liquidity-mining-begins-6e65932eaea9>

[McConaghy2018] Trent McConaghy, “Mission & Values for Ocean Protocol”, July 23, 2018, <https://blog.oceanprotocol.com/mission-values-for-ocean-protocol-aba998e95b8>

[McConaghy2019b] Trent McConaghy, “Datatokens 1: Data Custody”, Ocean Protocol Blog, FNov. 8, 2019, <https://blog.oceanprotocol.com/data-tokens-1-data-custody-1d0d5ae66d0c>

[McConaghy2019c] Trent McConaghy, “Datatokens 2: Non-Fungible, Fungible and Composable Datatokens”, Ocean Protocol Blog, Nov. 25, 2019, <https://blog.oceanprotocol.com/data-tokens-2-fungible-composable-54b6e0d28293>

[McConaghy2019d] Trent McConaghy, “Datatokens 3: Data and Decentralized Finance (Data \* DeFi)”, Ocean Protocol Blog, Dec. 3, 2019, <https://blog.oceanprotocol.com/data-tokens-3-data-and-decentralized-finance-data-defi-d5c9a6e578b7>

[McConaghy2020a] Trent McConaghy, “Radical Markets and the Data Economy”, Ocean Protocol Blog, Mar. 5, 2020, <https://blog.oceanprotocol.com/radical-markets-and-the-data-economy-4847c272f5>

[McConaghy2020c] Trent McConaghy, “How Does Ocean Compute-to-Data Relates to Other Privacy-Preserving Approaches?”, Ocean Protocol Blog, May 28, 2020, <https://blog.oceanprotocol.com/how-ocean-compute-to-data-relates-to-other-privacy-preserving-technology-b4e1c330483>

[McConaghy2020d] Trent McConaghy, “The Web3 Sustainability Loop. A system design for long-term growth of Web3 projects, with application to Ocean Protocol”, Ocean Protocol Blog, Sep. 1, 2020, <https://blog.oceanprotocol.com/the-web3-sustainability-loop-b2a4097a36e>

[McConaghy2020e] Trent McConaghy, “Ocean Datatokens: From Money Legos to Data Legos. How DeFi Helps Data, and Data Helps DeFi”, Ocean Protocol Blog, Sep. 8, 2020, <https://blog.oceanprotocol.com/ocean-datatokens-from-money-legos-to-data-legos-4f867cec1837>

[McConaghy2020f] Trent McConaghy, “Ocean Market: An Open-Source Community Marketplace for Data. Featuring OCEAN Staking, Automated Market Makers, and Initial Data Offerings,” Ocean Protocol Blog, Sep. 24, 2020, <https://blog.oceanprotocol.com/ocean-market-an-open-source-community-marketplace-for-data-4b99bedacdc3>

[McConaghy2020g] Trent McConaghy, “Ocean Token Model: System Dynamics for Near-Term Growth and Long-Term Sustainability,” Ocean Protocol Blog, Oct. 2, 2020, <https://blog.oceanprotocol.com/ocean-token-model-3e4e7af210f9>

[McConaghy2020h] Trent McConaghy, “Ocean Protocol V3 Architecture Overview: Simplicity and Interoperability via a Datatokens Core,” Ocean Protocol Blog, Oct. 12, 2020, <https://blog.oceanprotocol.com/ocean-protocol-v3-architecture-overview-9f2fab60f9a7>

[McConaghy2020i] Trent McConaghy, “On Selling Data in Ocean Market: Selling Your Proprietary Data, Others’ Proprietary Data, and Value-Adds to Open Data,” Ocean Protocol Blog, Oct. 19, 2020, <https://blog.oceanprotocol.com/on-selling-data-in-ocean-market-9afcfa1e6e43>

[McConaghy2020j] Trent McConaghy, “On Staking on Data in Ocean Market: Earning Opportunities — and Risks — in Ocean’s Community Marketplace,” Ocean Protocol Blog, Oct. 22, 2020, <https://blog.oceanprotocol.com/on-staking-on-data-in-ocean-market-3d8e09eb0a13>

[McConaghy2021a] Trent McConaghy, “NFTs & IP 1: Practical Connections of ERC721 with Intellectual Property”, Ocean Protocol Blog, Jun. 3, 2021, <https://blog.oceanprotocol.com/nfts-ip-1-practical-connections-of-erc721-with-intellectual-property-dc216aaf005d>

[McConaghy2021b] Trent McConaghy, “NFTs & IP 2: Leveraging ERC20 Fungibility”, Ocean Protocol Blog, Jun. 4, 2021, <https://blog.oceanprotocol.com/nfts-ip-2-leveraging-erc20-fungibility-bcee162290e3>

[McConaghy2021c] Trent McConaghy, “NFTs & IP 3: Combining ERC721 & ERC20”, Ocean Protocol Blog, Jun. 5, 2021, <https://blog.oceanprotocol.com/nfts-ip-3-combining-erc721-erc20-b69ea659115e>

[McConaghy2022a] Trent McConaghy, “Ocean Data Farming is Launching”, Ocean Protocol Blog, Jun. 9, 2022, <https://blog.oceanprotocol.com/ocean-data-farming-is-launching-f633532d831c>

[McConaghy2022b] Trent McConaghy, “Introducing veOCEAN”, Ocean Protocol Blog, Jul. 12, 2022, <https://blog.oceanprotocol.com/introducing-veocean-c5f416c1f9a0>

[McDonald2020] Mike McDonald, “Building Liquidity into Token Distribution”, Mar. 4, 2020, <https://medium.com/balancer-protocol/building-liquidity-into-token-distribution-a49d4286e0d4>

[Melon2020] “Melon Protocol”, homepage, last accessed Aug. 19, 2020, <https://melonprotocol.com/>

[MOBI2020] “MOBI - Mobility Open Blockchain Initiative”. Homepage. Last accessed Aug. 20, 2020, <https://dlt.mobi/>

[Monegro2020] Joel Monegro, “Proof of Liquidity”, Placeholder blog, May 22, 2020, <https://www.placeholder.vc/blog/2020/5/22/proof-of-liquidity>

[Napheys2020] Alex Napheys, “Introducing OceanDAO: A grants DAO curated by the Ocean community, for growth & long-term sustainability”, Ocean Protocol Blog, Nov. 30, 2020, <https://blog.oceanprotocol.com/introducing-oceandao-efe9ed681b84>

[NewOrder2021] “Introducing H2O - Our Latest Incubation Project”, New Order blog, Nov 30, 2021, <https://mirror.xyz/0x13c5432CfC12bA1f32B6090d1a09cf0Efe9C95Bd/LSazD5EFTpVptEGjeXofNRZyOI0vifSMfrvdvWloVf8>

[OceanBlog2020b] Ocean Protocol Team, “Ocean Protocol delivers Proof-of-Concept for Daimler AG in collaboration with Daimler South East Asia”, July 7, 2020, <https://blog.oceanprotocol.com/ocean-protocol-delivers-proof-of-concept-for-daimler-ag-in-collaboration-with-daimler-south-east-564aa7d959ca>

[OceanBlog\_Roadmap2019] “Ocean Roadmap || Setting sail to 2021”, Ocean Protocol Blog, Feb. 13, 2019, <https://blog.oceanprotocol.com/ocean-roadmap-setting-sail-to-2021-70c2545547a7>

[OceanBlog\_Roadmap2020] “Ocean Product Update || 2020”, Ocean Protocol Blog, Mar. 19, 2020, <https://blog.oceanprotocol.com/ocean-product-update-2020-f3ae281806dc>

[OceanContracts2022] "Ocean Contracts", Ocean Protocol GitHub, last accessed Feb. 24, 2022, <https://github.com/oceanprotocol/ocean-contracts>

[OceanDocs2022] "Ocean Protocol Software Documentation", last accessed Feb. 24, 2022, <https://docs.oceanprotocol.com/>

[OceanDocsNetworks2022] "Supported Networks", Ocean Protocol documentation, last accessed Sep. 1, 2022, <https://docs.oceanprotocol.com/core-concepts/networks>

[OceanDocsPermissions2022] "Fine-Grained Permissions", Ocean Protocol documentation, last accessed Sep. 1te, 2022, <https://docs.oceanprotocol.com/core-concepts/did-ddo#rules-for-did-and-ddo>

[OceanDAOwiki] "OceanDAO Wiki", homepage, last accessed Feb. 24, 2022, <https://github.com/oceanprotocol/oceandao/wiki>

[OceanLibJs2022] "Ocean-lib: JavaScript", Ocean Protocol GitHub, last accessed Feb. 24, 2022, <https://github.com/oceanprotocol/ocean-lib-js>

[OceanLibPy2022] "Ocean-lib: Python", Ocean Protocol GitHub, last accessed Feb. 24, 2022, <https://github.com/oceanprotocol/ocean-lib-py>

[OceanMarket2022] "Ocean Market", Ocean Protocol GitHub, last accessed Feb. 24, 2022, <https://github.com/oceanprotocol/market>

[OceanPurgatory2022] "Ocean Market Purgatory Process", Ocean Protocol GitHub, last accessed Feb. 23, 2022, <https://github.com/oceanprotocol/list-purgatory/tree/main/policies>

[OpenMined2020] "OpenMined", homepage, last accessed Aug. 20, 2020, <https://www.openmined.org/>

[Oauth2020] Oauth Team, "Oauth 2.0 Spec", Oauth Website, last accessed Aug. 19, 2020, <https://oauth.net/2/>

[Patel2020] Manan Patel, "Technical Guide to Ocean Compute-to-Data", Ocean Protocol Blog, May 19, 2020, <https://blog.oceanprotocol.com/v2-ocean-compute-to-data-guide-9a3491034b64>

[RBAC2022] "Role-based access control", Wikipedia, last accessed Feb. 23, 2022, [https://en.wikipedia.org/wiki/Role-based\\_access\\_control](https://en.wikipedia.org/wiki/Role-based_access_control)

[Rouviere2017] Simon de la Rouviere, "Introducing Curation Markets: Trade Popularity of Memes & Information (with code)!", Medium, May 22, 2017, <https://medium.com/@simondlr/introducing-curation-markets-trade-popularity-of-memes-information-with-code-70bf6fed9881>

[Samani2018] Kyle Samani, "New Models for Utility Tokens", Feb. 13, 2018, <https://multicoin.capital/2018/02/13/new-models-utility-tokens/>

[Solidity2020] "Solidity", Wikipedia, last accessed Aug. 21, 2020, <https://en.wikipedia.org/wiki/Solidity>

[Set2020] "TokenSets", homepage, last accessed Aug. 19, 2020, <https://www.tokensets.com/>

[Shopify2020] "Shopify", homepage, last accessed Aug. 21, 2020, <https://shopify.com>

[Simler2018] Kevin Simler and Robin Hanson, "The Elephant in the Brain: Hidden Motives in Everyday Life", Oxford University Press, Jan. 2018, <https://www.amazon.com/Elephant-Brain-Hidden-Motives-Everyday/dp/0190495995>

[Simmons2017] Andrew Simmons, "Prediction markets: How betting markets can be used to create a robust AI", Sept. 13, 2017, <http://dstil.ghost.io/prediction-markets/amp/>

[SWP2020] "Saskatchewan Wheat Pool", Wikipedia, last accessed Aug. 21, 2020, [https://en.wikipedia.org/wiki/Saskatchewan\\_Wheat\\_Pool](https://en.wikipedia.org/wiki/Saskatchewan_Wheat_Pool)

[TokenSPICE2020] "TokenSPICE: Token simulator using agent-based modeling", Ocean Protocol GitHub organization, open-sourced in Nov 2020, <https://github.com/oceanprotocol/tokenspice>

[Turley2020] Cooper Turley, "Moloch Evolved: V2 Primer", Mar. 27, 2020, <https://medium.com/raid-guild/moloch-evolved-v2-primer-25c9cdeab455>

[Uma2020] Uma Team, "UMA Initial Uniswap Listing", Apr. 22, 2020, <https://medium.com/uma-project/umas-initial-uniswap-listing-afa7b6f6a330>

[Unisocks2020] "Unisocks", homepage, <https://unisocks.exchange/>

[Vogelsteller2015] Fabian Vogelsteller and Vitalik Buterin, "ERC-20 Token Standard", Nov. 19, 2015, <https://github.com/ethereum/EIPs/blob/master/EIPS/eip-20.md>

[Vogelsteller2020] Fabian Vogelsteller, "rICO - The Reversible ICO", Lukso Blog, Apr. 13, 2020, <https://medium.com/lukso/rico-the-reversible-ico-5392bf64318b>

[WIPO2022] "Summary of the Berne Convention for the Protection of Literary and Artistic Works (1986)", World Intellectual Property Organization (WIPO), last accessed Feb. 24, 2022, [https://www.wipo.int/treaties/en/ip/berne/summary\\_berne.html](https://www.wipo.int/treaties/en/ip/berne/summary_berne.html)

[W3C2019] W3C, “Verifiable Credentials Data Model 1.0”, W3C Recommendation, November 2019, <https://www.w3.org/TR/vc-data-model/>

[XRoad2020] “X-Road”, Wikipedia, last accessed Aug. 19, 2020, <https://en.wikipedia.org/wiki/X-Road>

## **Acknowledgements**

Thanks very much to the following people for reviewing the paper and providing feedback: Simon de la Rouviere, Julien Thevenard, Bruce Pon, Alex Coseru, Monica Botez, Sarah Vallon, David Holtzman, Simon Mezgec, Robin Lehmann, Laurent Rochat, Gary Latta, Andreas Fauler. Thanks to my excellent colleagues at Ocean Protocol for the collaboration in building towards this. Finally, thanks to the broader Ocean community for their ongoing support. --Trent