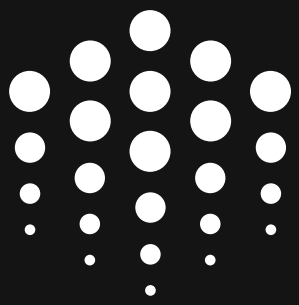


ocean

A Decentralized Data Exchange Protocol to Unlock Data for Artificial Intelligence

Reference Marketplace Framework

Ocean Protocol Foundation Ltd

February 2019

A joint project of

BIGCHAIN DB DEX

Abstract

This document presents a summary of the core marketplace attributes and components required to facilitate the successful deployment of the decentralized data exchange protocol and network called Ocean Protocol. It is complementary to the technical white paper for Ocean Protocol.

Modern society runs on data. Modern artificial intelligence extracts value from that data. However, the power of both data and AI is siloed. The goal of Ocean Protocol is to liberate data and open it up to AI, thereby distributing the power of data and AI. This liberation will be driven by asset tokenization propelled by blockchain.

ocean

Table of Contents

1. Introduction

OCEAN PROTOCOL ("OCEAN") IS A DECENTRALIZED DATA EXCHANGE PROTOCOL AND NETWORK THAT INCENTIVIZES THE PUBLISHING OF DATA FOR USE IN THE TRAINING OF ARTIFICIAL INTELLIGENCE (AI) MODELS.

The network facilitates the distribution and consumption of data and associated services in a safe, secure, and transparent manner. Ocean provides the mechanism for storing every asset's metadata including links to the asset itself, asset ownership, and associated IP licensing information.

On top of the protocol sit marketplaces that access and serve the underlying assets. Each marketplace acts as the last mile in connecting data providers with consumers. Ocean incentivizes uploading of high-quality data, including data intended for use in public data commons, as well as the services associated with consuming the data. Control of assets within the Ocean Protocol network is provided to the respective rights holder, with first-class privacy measures baked in. It also provides programmable market mechanics, making fair, yet flexible pricing easy. Additionally, Ocean is designed for industrial-scale usage.

Look no further than the government of the United Kingdom for the rationale driving Ocean. According to the report on *Growing the Artificial Intelligence Industry in the UK*, released jointly by the Department for Digital, Culture, Media & Sport and Department for Business, Energy & Industrial Strategy on October 15th, 2017, the UK "...could add an additional USD $814 billion (£630bn) to the UK economy by 2035, increasing the annual growth rate of GVA from 2.5 to 3.9%."

> *[However,] to continue developing and applying AI, the UK will need to increase ease of access to data in a wider range of sectors. This Review recommends:*
>
> - *Development of data trusts, to improve trust and ease around sharing data*

- *Making more research data machine readable*
- *Supporting text and data mining as a standard and essential tool for research.*[1]

There is discernible motivation for adopting AI, as highlighted above. It is also apparent that impeding AI growth will have adverse effects on economies, and likely on society as well. We find ourselves at an inflection point, and it is our strong belief that Ocean Protocol provides a clear path forward.

This document introduces Ocean Protocol's marketplace requirements. It is complementary to other documentation, including the technical whitepaper for Ocean Protocol.

1.1. Key Drivers

The primary goal of the Ocean network is to create a global supply chain of data and associated services for consumption by AI. This data will be of two types: "commons", or free assets, and in the future, priced assets. The assets themselves can be provided in raw form, or "cleansed" and modelled. Marketplaces will facilitate access to the assets made available by Ocean Protocol, and cater to the specific needs of their consumer base.

Critical to each marketplace is ensuring provenance. This virtual paper trail is immutable and inherent to the network. It creates trust in the ecosystem for all participants. Ocean network allows asset providers to control who (or what) accesses their assets, as well as how and where the assets are being used. In essence, Ocean acts as a data DNS with associated access control and service level agreement (SLA) obligations.

The Ocean network facilitates curation by providing economic signals at the network level. Marketplaces can use these signals to aid discovery on their own platforms. Curation is a critical element to any industrial-scale data and services platform, providing indicators for robustness, reliability, and relevance. This capability will manifest in numerous patterns within Ocean, like data mashing,

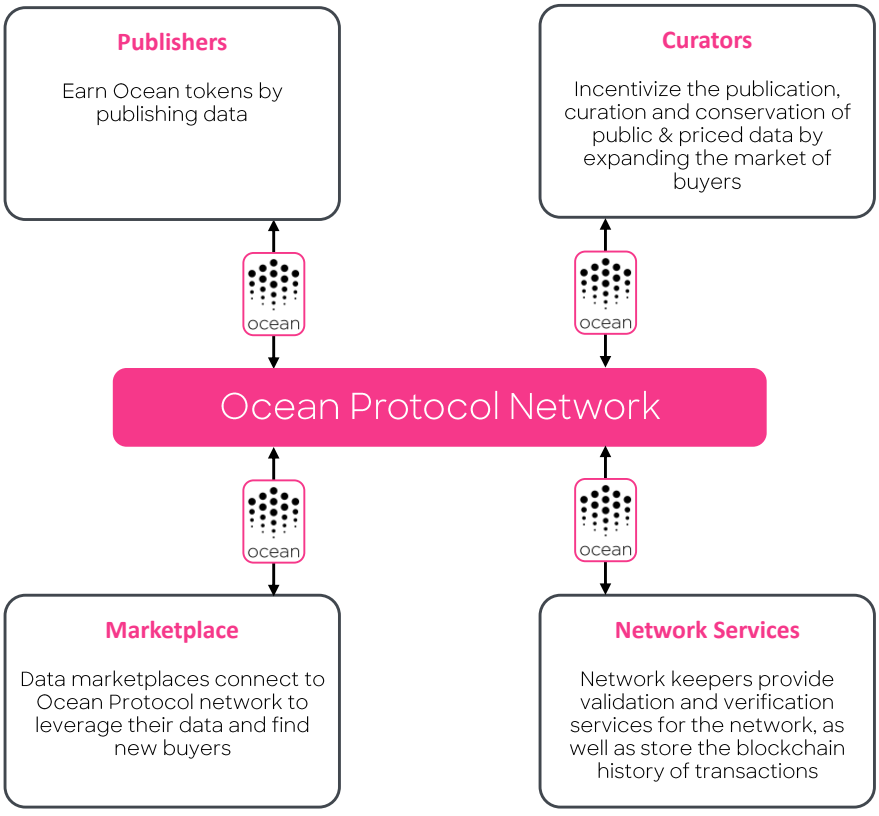
[1] https://www.uk.gov

semantic data mapping, and data integration. Orchestrated together, this functionality will provide a compelling value-added service to the network and marketplaces.

In addition, Ocean provides a Data Science layer on top of the protocol. Data Scientists, AI practitioners, and analysts can access tools like Jupyter notebooks that are directly linked with Ocean. A Data Scientist can thus search, access, and run analysis using a familiar workbench and tools directly against Ocean assets. Additionally, they can register their algorithmic assets with Ocean, and open up new collaborative opportunities with other experts within the ecosystem.

1.2. Ecosystem: Marketplaces, Network

The following diagram illustrates the relationship between asset providers, asset consumers, marketplaces, and the Ocean network. This document focuses on data marketplaces, and how they interface with other actors in the ecosystem.



Here are the entities in the ecosystem:

- **Ocean Ecosystem** - A collection of marketplaces, the Ocean network, assets and services, and related actors.
- **Ocean Data Marketplace (DM)** - A data marketplace designed to use the services of Ocean network to simplify its implementation, and for increased liquidity of data supply and demand.
- **Reference (Ocean) Data Marketplace** - A data marketplace with open-source code and (as much as possible) open-source legals, that other would-be marketplaces can use to get started quickly. The first version is a Commons Marketplace enabling discovery and consumption of open datasets.
- **Ocean Network** - The decentralized network that incentivizes for the supply of relevant data and services, with a user registry, access control, and other mechanisms to mitigate bad behavior. The network incentivizes/aligns interests using Ocean token (OCEAN) and uses metadata stores for the identification of asset details.
- **Publishers** - The actors in the ecosystem that register assets for discovery and consumption. In return for making the assets available, publishers will receive Ocean token (OCEAN). These actors can publish assets independently, or they can leverage existing marketplaces in order to further promote their assets.
- **Curators** - The actors in the ecosystem that signal relevant assets. These actors discover assets, both on and off-chain, and promote the adoption of assets through signaling mechanisms (i.e. up/down voting on relevance of an asset for a specific use case). Curators are in large part responsible for the growth of the ecosystem.
- **Ocean Protocol** - The protocol spoken by Keepers (nodes) in the Ocean network. A protocol is a specification of how machines talk to each other to accomplish the goals of the network. Technically there could be many networks all speaking the same protocol. However, we envision one large network (Ocean network) as it will aid with market liquidity and align incentives around one token (OCEAN).

A later section describes the other stakeholders in the ecosystem.

2. Ocean Marketplace Overview

THE FOLLOWING IS A SUMMARY OF THE KEY TECHNICAL ATTRIBUTES REQUIRED FOR A FULLY FUNCTIONING MARKETPLACE WITHIN THE OCEAN DATA EXCHANGE PROTOCOL.

Further details of Ocean's underlying technology stack are covered in the technical white paper.

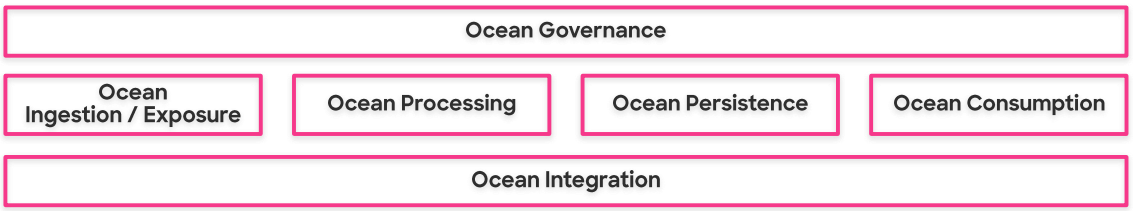
2.1. Reference Marketplace Capabilities

Marketplaces running on top of Ocean Network will support the following core capabilities:

1. **Data Exposure/Ingestion** - Data assets are exposed by data providers. These assets can be raw data with little to no modelling, or fully transformed data models, similar to what would be found in enterprise data warehouses. These assets can reside within the network, as is the case with free public data, or outside the network behind firewalls. In the case of Ocean Commons, actors can add open datasets by pasting the URL of the asset, and publishing the metadata. Ocean incentivizes actors to discover and publish these open datasets with rewards.
2. **Data Processing** - Data processing provides the compute mechanisms required to cleanse, transform, and analyze exposed data. Ocean Protocol's processing functionality provides data curators with the ability to normalize exposed data in order to create new assets, while keeping track of source or background IP. This capability also provides the means for deploying and collectively developing AI algorithms. Processing can be provided on-premise behind firewalls when required, by data

marketplaces, or by registered data processors within the network. In all cases, the compute moves to the data.

3. **Data Persistence** - Data persistence provides the mechanisms for storing post-processing result sets. These mechanism can provide simple distributed blob or file storage, similar to HDFS, or decentralized mechanisms like IPFS[2], Storj[3], Swarm[4], etc. It can also be MOLAP or ROLAP data stores for analytic consumption, in-memory persistence for low-latency data access, tuple or document stores for scalable operational data storage, highly indexed data storage optimized for search, etc.
4. **Consumption** - Consumption provides the means for end-users/consumers to leverage the underlying assets. This mechanism is generally provided by marketplaces providing an interface to the data, whether B2C, B2B, M2M, etc.
5. **Integration** - The integration mechanism provides secure, end-to-end access to the network's assets while enforcing authentication and authorization protocols. Reference marketplace integration capabilities manifest as APIs and microservices deployed and maintained by registered integration providers.
6. **Governance** - Governance is a first-class citizen in marketplaces for Ocean. Asset provenance is baked in by establishing an immutable record of all transactions within the network via blockchain. Also, Ocean network's curation markets enable the creation of standardized data dictionaries for Master Data Management (MDM).
7. **Utility** - Ocean network itself is a utility as it provides the basic infrastructure substrates for a public service. These substrates include the means for transacting within the network using Ocean tokens, and marketplace protocols that orchestrate Ocean's capabilities.[5]



[2] https://ipfs.io/
[3] https://storj.io/
[4] http://swarm-gateways.net/bzz:/theswarm.eth/
[5] https://news.21.co/thoughts-on-tokens-436109aabcbe

2.2. Ocean Tokens

Ocean tokens are the means of transacting within the ecosystem, and because the tokens can be exchanged to procure network services, they are treated as utility tokens.[6] Additionally, Ocean's blockchain technology removes the possibility of infinite reproducibility in digital assets like data and algorithms.[7] As such, it can be confirmed that each unit of value is transferred only once, solving the long-standing problem of double spending.[8]

Ocean facilitates the tokenization of all assets. Tokens can be acquired through purchase via external third party offerings, or by offering a value-added service (i.e. providing data and "keeping") within the ecosystem. A key network contributor is the data marketplace that functions on top of the network. The role of the marketplace is varied, as it can directly expose Ocean tokens to its users, or it can let its users operate in some fiat currency with the network facilitating exchange.

Take, for example, a data provider. As a network asset contributor, a data provider can make data available in exchange for tokens (typically via a marketplace). To access this data asset, a data consumer only needs to provide the provider with the requisite number of tokens. The handshake between counterparties is a simple mechanism handled by Ocean's underpinning Service Execution Agreements, or SEAs (for more details on SEAs, please see the technical whitepaper).

With tokenization, Ocean offers a common mechanism of exchange to reduce the friction generally associated with data sharing and procurement.

---

[6] https://news.21.co/thoughts-on-tokens-436109aabcbe
[7] https://en.wikipedia.org/wiki/Reproduction_(economics)
[8] https://en.wikipedia.org/wiki/Double-spending

2.3. Key Ecosystem Stakeholders & Network Contributors

There are eight key stakeholders in the ecosystem: Asset Providers, Asset Consumers, Marketplaces, Data Scientists, Data Service Providers, Referrers, Network Keepers, and Regulators. Each plays a unique and critical role:

Asset Providers

Asset Providers are the core actors to the Ocean ecosystem. They provide the network with data and associated services in exchange for tokens, or for the data commons. The data assets provided may be raw data files, blobs, structured, semi-structured, unstructured, etc. The data may be heavily modelled and available for usage as MOLAP or ROLAP data, or completely un-modelled and available via distributed file stores like HDFS or IPFS. Asset Providers can be broken down into the following subsets:

- **Asset Owner** - Asset Owners are the original proprietor and purveyor of the asset. They legally own the data intellectual property (IP) or service and can facilitate usage of their assets when compliant with regulations.
- **Asset Custodian** - Asset Custodians hold assets, such as data, on behalf of their customers, as well as maintain the value of the assets in compliance with regulations. They do this by validating assets against benchmarks for usability, accuracy, and relevance. They are also responsible for creating and maintaining the metadata mappings for any data asset.

Asset Consumers

Asset Consumers are the primary users and beneficiaries of Ocean's assets. Ocean consumption is open to all, and will be made up of individuals, start-ups, small to medium sized companies, and large-scale multinational enterprises and governments. As stated previously, it is Ocean Protocol's goal to open up access to an extensive array of varied data and services for use by AI. As such, Ocean is perfectly suited to meet the needs of AI specialists, Data Scientists, Big Data Engineers, and Business Intelligence professionals.

Data Scientists

Data Scientists (and AI researchers/engineers) are a crucial part of the Ocean Protocol ecosystem. They are the primary consumers of data and services as they require access to datasets and compute to train their models and run analyses. Given that one of the main goals of Ocean Protocol is to democratize access to data for AI training and Data Science, the needs of these key users are at center of the ecosystem. Ocean, especially in its earlier versions, is particularly geared towards simplifying the flow for data analysis. The ecosystem comes with a Data Science layer called Manta Ray (more on this later) with a set of tools benefitting these users. For example, a Data Scientist can directly launch Jupyter notebooks and synchronize with Ocean Protocol to work on the datasets they have gained access to.

Marketplaces

Marketplaces play the role of intermediary between Asset Producers and Asset Consumers. Each marketplace provides access to a set of assets, exposing the assets to its subscribed consumers. The marketplaces also facilitate the transmission of assets between counterparties in a seamless fashion.

Data Service Providers

Data Service Providers are a vital part of the ecosystem, as they offer different data services, which can benefit asset consumers, asset providers, or marketplaces. Similar to the data providers, they can receive tokens in return of their services. In future releases, users will be able to stake tokens on these data services as way of curating the relevant services to the community. The types of services that exist on Ocean Protocol can be numerous and diverse. For example, there can be data mashers, data labelling/tagging services, data cleaning services, data compute services, analytics, etc. Ocean Protocol provides some reference implementations for these services, and will expand capabilities over time by deploying home-grown solutions or by integrating with partners. However, we ultimately expect the community to deploy the broadest range of data services.

Referrers

Referrers promote the use of Ocean to Asset Providers, and facilitate the linking of Asset Consumers to published assets. Consequently, the responsibility of identifying valuable assets and their corresponding purveyors is that of Ocean's Referrers. This key role manifests itself through the development of marketplaces, from which assets are procured from providers and exposed to consumers.

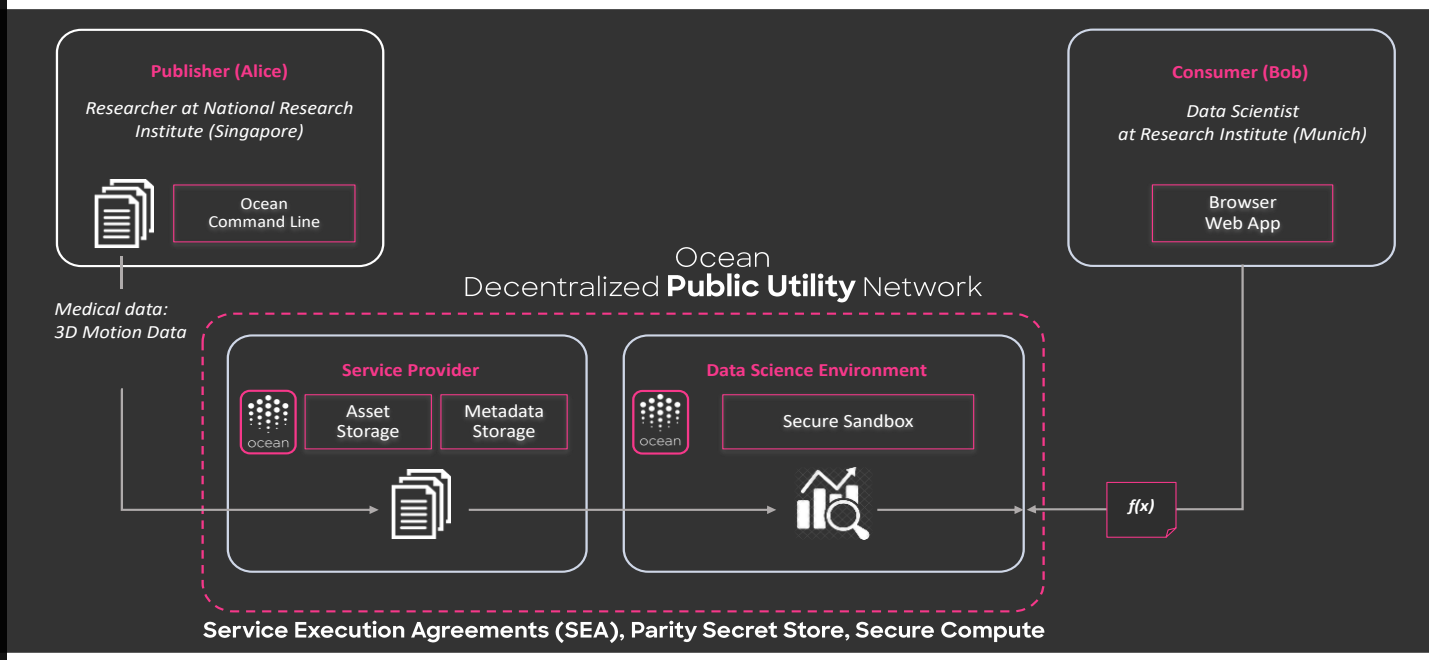
Network Keepers

Network Keepers provide and manage the orchestration of Ocean's critical substrate functionality. Keepers run as nodes within the network and provide core provenance and attribution capabilities. It is envisioned that Data Providers, Marketplaces, and Referrers will eventually make up a large proportion of Network Keepers. However, these services can also be provided by third party contributors with expertise in specific functional areas, like Data Integration or Data and Platform Auditing.

For more details on Network Keepers, please see the technical whitepaper.

Regulators

While this may be contrary to popular opinion, regulators are critical to provide guidance for the protocol and network. Ocean's use of blockchain does not absolve contributors of their requirement to protect data assets to the utmost. Inclusion of all vested parties is critical to Ocean's success. In fact, working with regulators and auditors to satisfy compliance helps to reduce overall friction within the network, as it removes contributory reluctance, as well as impediments to consumption. The added benefit is that these safeguards can be intrinsically tokenized within the protocol, adding even more impetus to play by the rules.



Most contributors will access the Ocean ecosystem via data marketplaces that will be built on top of the Ocean network. Others, like regulators, will interface with the Ocean Protocol Foundation directly. Any new services in the ecosystem will have access points at the network level and via the marketplaces.

2.4. Client "*Squid*" Libraries

Ocean Protocol is based on three layers. The bottom layer provides the smart contracts via the blockchain side of the network. The top layer includes the marketplaces and Data Science toolsets. Connecting these two layers are the client libraries, referred to as the *Squid Libraries* in Ocean. These libraries abstract the complexity of the smart contracts layer by providing easier interactions to the core functions of Ocean Protocol: publishing, searching, and consuming assets. They are written in:

- JavaScript for use in front-end development;
- Python for use by Data Scientists, and other AI and advanced analytic developers; and,
- Java for data engineering development.

The Squid Libraries represent the glue of Ocean Protocol and the easiest way for developers to contribute to the ecosystem by building new features, tools, and services.

2.5. Data Science Toolset – Manta Ray

In a traditional, two-sided marketplace there are buyers and sellers. In the case of Ocean, there are asset providers and asset consumers. Specifically, Ocean caters to Data Scientists and AI practitioners. Consequently, a key focus is to incorporate Ocean into the standard Data Science workbench. This inclusion is handled via Manta Ray.

Manta Ray makes the experience for Data Scientists more seamless by providing the toolsets necessary for them to complete their tasks. It leverages the underlying Ocean API libraries that abstract the smart contracts layer. As mentioned earlier, these libraries provide the means to search, publish, and consume assets. Manta Ray exposes these functions within tools familiar to Data Science practitioners. Specifically, Ocean Protocol has configured JupyterHub to spawn JupyterLab instances in the consumer's browser. Leveraging this popular tool, users now have the ability to access relevant assets like data or algorithms exposed via Ocean.

Ocean Protocol is offering more than just the traditional data exchange marketplace framework, but a real set of tooling for Data Scientists.

2.6. The Marketplace Honeypot Solution

One of the risks with data marketplaces is the unauthorized accessing of data assets by malicious actors. In the case of a classic data marketplace, data is consolidated into a single repository, making them targets of attack. And because of the consolidated nature of the deployment, once the authorization layer has been compromised, malicious actors are basically free to access, steal, and manipulate data information at will.

Ocean, however, takes a more “belt and suspenders” approach to this problem, implementing multiple independent authentication and authorization mechanisms. Ocean marketplaces have two levels of defense against malicious activity can have two layers of defense against bad actors. An authentication and an authorization one.

Firstly, a marketplace on Ocean can have its own set of governance mechanisms and can require different conditions be fulfilled for a user to access the marketplace. As an example, imagine a genomics marketplace. The governance model could state that only verified genomics researchers (validated via offline verification for example) can access the marketplace. Alternatively, the marketplace could rely on a vetting system, or a referral one. A malicious actor must crack this first layer to access the marketplace and view the different assets available via the marketplace. However, because the marketplace does not physically consolidate the data, but only references it, the attack surface is limited. (Please note: For the Commons Marketplace, there is no strong restriction on users accessing the marketplace, as it’s for open and free datasets.)

If a malicious actor does breach a marketplace’s initial line of defense, they next need to target the individual assets that the marketplace promotes. For this, the malicious actor must breach the authorization mechanism for each specific asset published to the marketplace. In the case of the genomics marketplace example, each marketplace dataset could be sitting behind a unique firewall, each on-premise within the providers’ data center. In order for an malicious actor to access all assets, the actor would need to breach each providers security mechanism. Additionally, each access request made by the malicious actor would be committed on-chain, leaving a trail of malicious activity behind.

Consequently, the honeypot effect that takes place in traditional data marketplaces is significantly reduced due to Ocean’s network architecture.

2.7. Data Governance

Data Governance is critical to the successful operation of any data platform. As such, Data Governance is provided first-class citizenship within the Ocean

ecosystem. At the core of this is provenance, which manifests itself through the immutable nature of transactions within Ocean's blockchain. Any transaction event that occurs within Ocean is recorded. Creating a virtual breadcrumb trail of all transactions within the network stack makes establishing provenance and auditability relatively easy.

With this in mind, Data Provenance has been a core focus of BigchainDB ("BDB"), a scalable blockchain database provider, since its inception, and remains so after solution adoption by 40+ corporates. By providing connectors/API plug-ins to BDB nodes for all integration points within the IPDB network, users can track data usage throughout the network. In the case of regulatory compliance, this capability is especially beneficial as compliance issues often subside so long as verifiable audit and provenance can be established. BDB can be used as metadata store in the context of Ocean (more on this below).

Furthering the Data Governance capabilities, trusted curated registries work in tandem with staking (more on this later) to facilitate the deployment of best-of-breed governance policies and standards. These registries will provide an adoption mechanism for standardized Master Data Management (MDM) policies, and associated Data Dictionaries, potentially across entire domains (or even across domains). Applying these policies could be as simple as subscribing to the top registry entry, and enforcing the associated policies and framework to an existing data asset.

2.8. Metadata Store

Multiple marketplaces are on Ocean, some of them specialized in certain verticals, or domain spaces with specific types of data. However, they do have a lot of the design in common, including the need for a metadata store.

When an asset is published via Ocean Protocol, its metadata needs to be stored so that it becomes easily resolvable and searchable. Additionally, the integrity of the asset may also require verification. Currently, several options exist to facilitate these features. Current support for metadata store deployments exists for BigchainDB, MongoDB, and ElasticSearch. What type of backend is

used, however, is ultimately up to the marketplace itself. This design allows for fast and efficient search, and easier integration.

In the first iteration of Ocean's data marketplace, the Ocean Commons Marketplace, users can publish open datasets by filling in pertinent asset metadata via the marketplace interface, ensuring a robust and large marketplace.

2.9. On-Chain Access Control

Ocean Protocol uses an access control mechanism to authorize users on-chain when connecting a publisher/provider and a consumer. This guarantees that only the verified user can access the asset. As it is on-chain, it means that the access policies are embedded in the smart contracts run by Ocean Protocol. In this case, Ocean is using Parity Secret Store which allows decryption of data once a user is authorized on-chain. The Secret store is a set of nodes running secure authorization software, each saving a part of the decryption key. Once consensus is reached, the decryption key is released to the authorized user. This mechanism ensures security and traceability.

2.10. Ocean Marketplace Deployment Strategy

Ocean's marketplaces will act as Grand Bazaars for data and services, enticing consumers with their alluring assets, while at the same time attracting asset providers to the network because of access to a broad consumer base. Each marketplace can cater to specific domains by providing data relevant to that domain only, or appeal to a broader consumer base by providing value-added services like access to cross-domain data assets and mash-ups.

In the first release version of the Ocean POA, the reference marketplace is an Ocean Commons Marketplace, where consumers can search and access open datasets that have been published by providers. This first step will ultimately lead to paid data assets marketplaces.

Initially, the inclination will be for marketplaces to manifest as holistic, end-to-end solutions that include the means to ingest, process, persist, consume, and govern data. Gradually, however, network effects will take over, and more suitable providers with distinct capabilities and expertise will emerge. This means that over time, providers of specific capabilities (e.g. in-memory persistence) will materialize to compete with generic marketplace offerings. As a result, marketplaces will naturally begin to decompose in favor of a modular approach. By doing so, best-in-class marketplaces will be in constant flux as they select capabilities from the evolving best-in-class providers within the network.

This ethos may lead to a modularized, evolutionary design approach for marketplace solutions. Consequently, it could be in the marketplaces' best interests to link together the best end-to-end capabilities available from the network at that moment, in order to attract both providers and consumers to their offering. Eventually, the primary function of marketplace providers could be to act as referrers of both data and component capabilities.

2.11. Types of Data

Ocean will expose three primary types of data:

1. **Proprietary Data** - This is data that is controlled by a data provider/owner, and is generally unique to that provider/owner.
   Example: Proprietary autonomous vehicle data.
2. **Regulated Data** - This is data that is controlled jurisdictionally through regulation or other means. While the data may not be unique, its accessibility is limited generally due to privacy constraints.
   Example: Personal medical history data.
3. **Free or "Commons" Data** - This is data that is generally free or open for use. This type of data generally has limited restriction on its usage. This is the core category for the initial releases of Ocean Protocol. These datasets will be available via the Ocean Commons Marketplace for consumers to access and work on.
   Example: National Census data.

2.12. Pricing

It is envisioned that in the medium term, after the initial release (which will be for free and open data assets), Ocean will enlist three pricing schemes for data, depending on the data type and its fungibility or uniqueness:

1. **Free Data** - This data is open to all consumers with no restrictions. We want to encourage a growing data commons for the world. The token design elaborates on the incentive structure.
2. **Proprietary/Regulated Non-Free Fungible Data** - With data that is relatively universal but controlled, the pricing is easy(ish): just use an exchange. Exchanges are low friction and let the market determine the price. We plan to support data exchange functionality in the Ocean Protocol.
3. **Proprietary/Regulated Non-Free, Non-Fungible Data** - For data that is unique, pricing becomes more difficult. The price could simply be fixed. However, if priced too low, it's a lost revenue opportunity. And if priced too high, no one will buy it. Generally, market dynamics will eventually bring about pricing equilibrium, but this could take time. To address these concerns, we explored several pricing schemes and distilled them into three options: fixed price, auction, and royalties.

3. Engagement Model

THE FOLLOWING SECTION DESCRIBES HOW OCEAN WILL ENGAGE WITH BOTH ASSET PROVIDERS AND CONSUMERS.

3.1. Asset Providers

For any network to succeed, the right players must be activated at the right time. This is no different for Ocean. First and foremost, this initially means rigorous engagement with Asset Providers in order to prime the network. Without data and associated services, there is no Ocean.

The onboarding of providers will be relatively straightforward. Initially, all data will reside *in situ*, and be exposed to the network protocol via light-touch API's. To expose their assets, data providers will navigate to their marketplace portal of choice, and select the option to provide access to data. Next, they will register with the network, providing information about the data owner. The marketplace portal will allow the data provider to designate which assets should be exposed to the network, plus any consumption parameters that govern asset accessibility. All of the transactional information will submitted on-chain via the keeper nodes for provenance, and the assets will be exposed, in a controlled manner, to the consumer base.

3.2. Asset Consumers

Without the demand-side of the network, Ocean is unviable. Thus, proper engagement of Asset Consumers is critical. However, the timing of this activity is also essential. Too early, and consumers won't see the value of the network. Too late, and providers won't realize the return on their contribution. This *Goldilocks Dilemma* is exacerbated by Ocean's core target base of AI researchers and startups because AI's need massive amounts of data.

Auspiciously, we seem to have hit an inflection point, as AI adoption, along with the understanding of AI's inherent need for data, becomes more prevalent.[10] This understanding will assist in placating data providers who traditionally look for immediate return on investment. In Ocean's case, the potential upside for further AI advancement are too great to leave the ecosystem.

Onboarding Asset Consumers will be relatively simple. Like data providers, consumers will engage with their marketplace of choice. In certain instances, like for the Financial Services or Healthcare domains, onboarding to a marketplace may require KYC. For others, participation could be completely anonymous, in which trusted users access data via permissionless marketplaces. Depending on the marketplace and its associated providers, the available consumption mechanisms could include embedded dashboards and mash-up windows, as well as full access to data assets for download.

For AIs, the process will include additional steps, such as providing access to distributed sandbox environments, or the potential to encrypt AI algorithms and push them to encrypted, containerized data assets (i.e. via fitchain). This type of transaction is explained in more detail in the technical whitepaper.

3.3. Data Marketplaces

Data Marketplaces will provide the bridge between Data Providers and Data Consumers. Marketplaces for open commons assets will naturally gravitate towards the Ocean network in order to more easily access data and service. The Ocean ecosystem will also facilitate the creation of new marketplaces, as well as the amalgamation of existing ones. To expose data for consumption, each marketplace will define and deploy its own means of data exposure, potentially as competitive advantage.

Onboarding Data Marketplaces will first require making a request for inclusion into the network (whether registry driven or not). Once access to the network has been granted, the marketplace will request to connect to existing data assets, or expose net-new ones. The marketplace will also begin promoting the

[10] https://www.uk.gov

assets and attracting new consumers for onboarding, or facilitating the onboarding of existing users to the network.

3.4. Delivery Strategy - <Hello World>

All solutions have an inception point. The first release is centered around a Commons Marketplace. Data Scientists and AI practitioners are able to discover and consume open datasets via this marketplace. Opening access to data for the world is central at Ocean Protocol and the Commons Marketplace is the first step towards this goal.

In addition to the Ocean Commons Marketplace will be a set of parallel initiatives. Each of these initiatives targets a specific domain space to gain exposure to the broadest set of implementation requirements as possible. The goal is for these projects to influence the creation of a sandbox framework that can be extended beyond the initial contributors, and accelerate adoption of the Ocean network.

The table on the following page lists the themes and corresponding issues or opportunities for each of the initial four initiatives:

Theme	Partnerships	Issue/Opportunity
Mobility Including Trade Connectivity, Logistics, & Transportation	• MOBI https://dlt.mobi/	• Autonomous vehicles, route optimization, vehicle depreciation, environmental impact and road safety • Digital commerce and changing patterns in delivery to end customer (incl. B2B)
Healthcare Including Diagnostics & Therapy	• ConnectedLife https://connectedlife.io/ • Roche https://roche.com/ • J&J https://jnj.com/ • Aviva https://aviva.com/	• Consumerism of wellness and healthcare diagnostics, extending beyond the clinical setting and into the home • New measures of health, wellness and biomarkers enabling next generation of lifestyle recommendation engines
Consumer Products & Retail	• Point of Sale: NextBillion https://nextbillion.asia/ • Consumer Packaged Goods: Unilever https://unilever.com/ • E-commerce: DCI https://dci.ai/	• Optimization of physical and digital commerce – media, retail, incentives and payments
Utilities Including Energy & Water	• Verv https://verv.energy/	• Re-imagine and reduce energy and water consumption associated with activities of daily living and consumer products

3.5. Customer Engagement

Community engagement for a project of this nature is crucial. The larger that web of engagement, the greater the overall chance of success. Consequently, maintaining dialog and open lines of communication with the community are paramount to Ocean. Fostering trust requires openness and transparency in good times, and in bad. It mandates acting with decorum while always respecting all interested parties.

Of equal importance is the community's feedback loop into Ocean. Understanding the needs of stakeholders is critical in order to understand what's working, and what isn't. The community will be able to comment, advise, and provide ideas about the protocol to guide development. This is a symbiotic relationship—a healthy ecosystem demands an engaged community.

In concrete terms, this will involve establishing channels of communications through mediums like Telegram, Slack, and Twitter, as well as hosting AMA's with Ocean's technical and business teams. Additionally, updates will be provided via email to Ocean subscribers, and through blog posts on the Ocean website and medium.

By establishing trust between Ocean and our community, we will be able to build this ecosystem together, and finally unlock the true potential of AI and data.

4. Conclusion

This Reference Marketplace document is a guideline for what is required from Ocean's marketplaces, both in terms of capabilities, and in terms of functionality. The goal of Ocean is to unlock massive, disparate troves of data for consumption by AI, but to do so in an equitable and secure manner. Many of the concepts and principles presented here are established legacy designs, but many are new and novel and thus, subject to change and modification.

It is our hope that through engagement with the community, Ocean will be able to change the way we leverage data, and revolutionize the world with AI.

ocean

Ocean Protocol Foundation. A Non-Profit Foundation
www.oceanprotocol.com