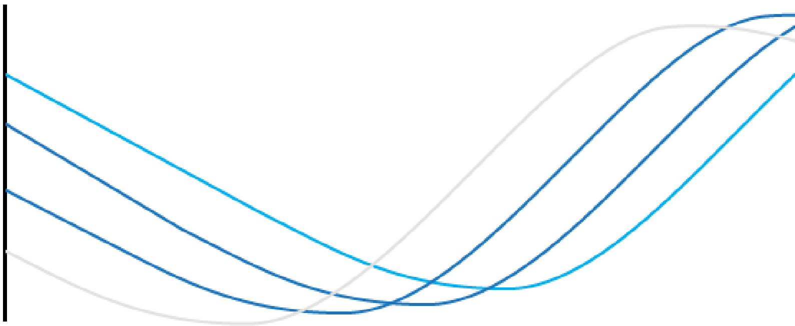
Ocean Protocol Use Case

Federated Learning



Accelerating progress in AI with Federated Learning on Ocean Protocol

Abstract

While so far, the advance of AI has been paid with the privacy of the people on a global scale, data is typically kept in silos available to a few companies only. With Federated Learning on Ocean Protocol, society can benefit from faster progress in AI without compromising its privacy and losing control over its data. Instead of data that is kept in silos and available to few, data access can be monetized by anyone without intermediaries in a decentralized fashion. On the other hand, data consumers can improve their machine learning and prediction models with data being unlocked by Ocean Protocol.

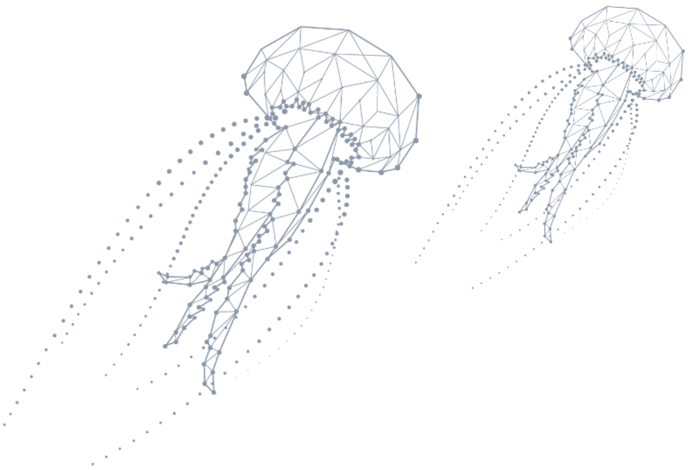
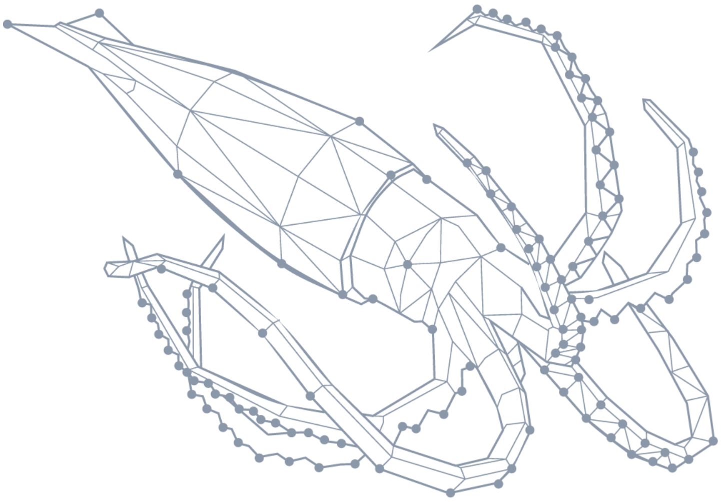


Table of Contents

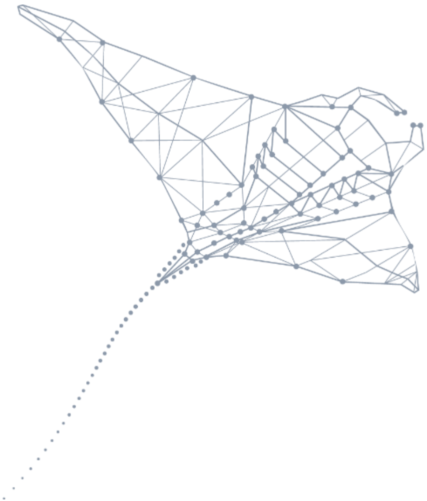


Introduction

Nowadays, the success of data-driven projects is highly dependent on the availability of high-quality, real-world data because the quality of a machine learning solutions heavily depends on the availability and quality of training data. While large amounts of valuable data are generated each year, this data remains unexchanged to a large extend because of concerns about data security, privacy, and trust.

Federated learning is a machine learning technique that allows training an AI model across multiple decentralized data stores (e.g., edge devices) holding local data samples without exchanging the data itself. It enables data consumers to build a global machine learning model without sharing or collecting data. If data isn't shared at all, critical issues of traditional centralized machine learning techniques where data is transferred to a server, such as data privacy, data security, data access control, and bandwidth, can be addressed more easily. With federated learning, machine learning on sensitive and private data becomes a viable solution in practice.

But we still need a trusted entity that governs the algorithms and the orchestration of the training across all those edge data stores. If the orchestrating entity has vulnerabilities or even acts maliciously, the sensitive and private data of all participants is at risk. This is where Web3 technology and Ocean Protocol can add tremendous value: The central entity becomes obsolete while adding properties such as transparency, provenance, and data sovereignty.



Ocean Protocol

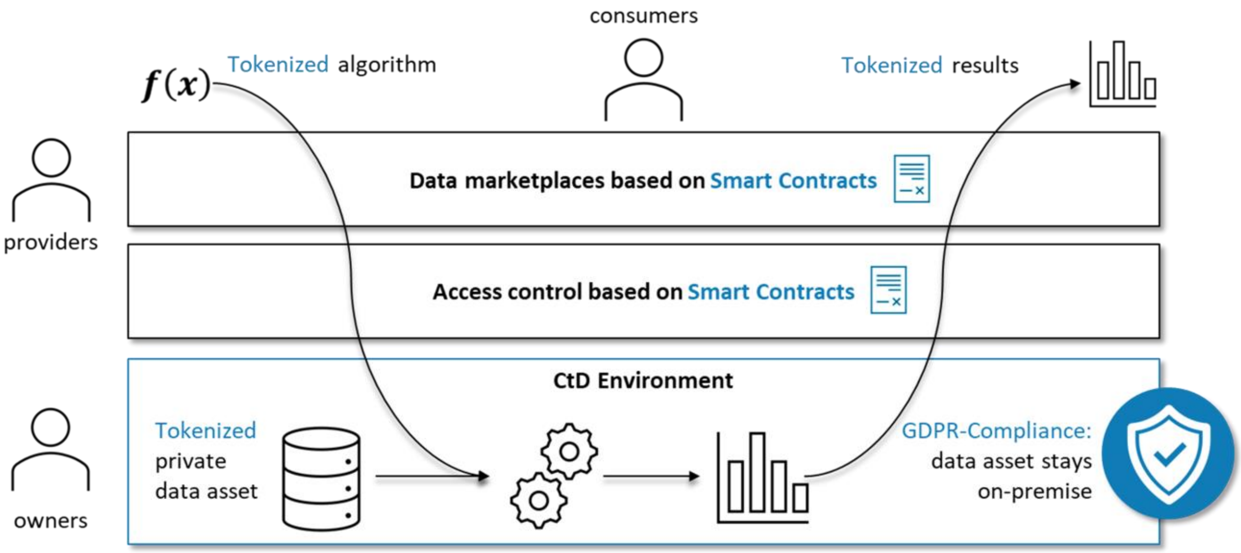
Ocean Protocol offers suitable technology to make decentralized Federated Learning a reality. Today, we aren't ready to use Federated Learning on Ocean Protocol in production. Still, all the crucial components necessary to implement a self-sovereign Federated Learning solution, running on a decentralized ledger governed and owned by no central entity, are already in place. It's a matter of when not if.

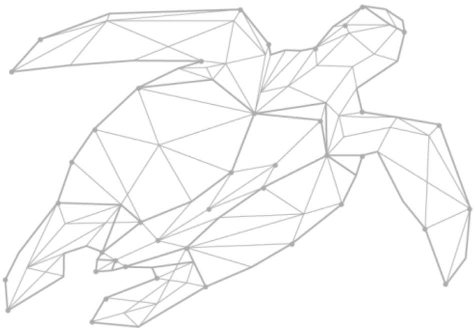
- Distributed Ledger Technology (DLT): The distributed ledger allows the maintenance of a global and append-only data structure by a set of mutually untrusted participants in a distributed environment. It holds a consensus of replicated, shared, and synchronized data, geographically distributed globally, based on a peer-to-peer network. There exists no central authority or administration. The characteristics of a distributed ledger, such as immutability and accountability, make it suitable to run a distributed data access control layer on top.
- Self-Sovereign Identity: SSIs give individuals control of their digital identities and addresses the difficulty of enabling trust in digital interactions. With SSI, trusted and centralized authorities are becoming obsolete.
- Smart Contracts: A computer program that automatically executes, controls, or documents events according to the terms of a digital contract is referred to as a smart contract. It is a collection of code (functions) and data (state) that resides on the distributed ledger. The program can contain the policy framework of the ecosystem, can be triggered by ledger transactions and is then executed on every node of the network automatically. Consequently, all nodes hold the state changes corresponding to the replicated execution of the program. Given that the smart contract layer inherits the features of the distributed ledger, nobody has central authority over the program execution. Smart contracts can be used to implement tokens, ownership, access control, taxes, voting, and similar logic on the blockchain. They guarantee and allow for immutability, transparency, and auditability in a permissionless and trustless environment.

Ocean Protocol combines those components (Distributed Ledgers + Self-Sovereign Identities + Smart Contracts) to enable a data access control layer that is fully decentralized and autonomous. Tokens represent data access rights on a distributed ledger, so-called "data tokens". Each data service has its own data token. Applied to the Federated Learning use case, each edge data storage and each algorithm have their data (access) token. Using the technology stack described before, we can access decentralized data. But how can we facilitate machine learning on that data?

Compute-to-Data

Compute-to-data allows for privacy-preserving and data-sovereign machine learning on remote data. The idea behind compute-to-data is to keep the data on-premises and to allow data consumers to run remote compute jobs on the data. Data owners keep full control as the data is never copied and exposed to the data consumer. Using Ocean Protocol, data consumers can train their machine learning models across many data sets of different data providers while ensuring compliance with data protection and privacy regulations. Furthermore, because Compute-to-Data on Ocean Protocol operates in a trustless and decentralized manner, the central authority that was needed to orchestrate the process of federated ML is no longer necessary.

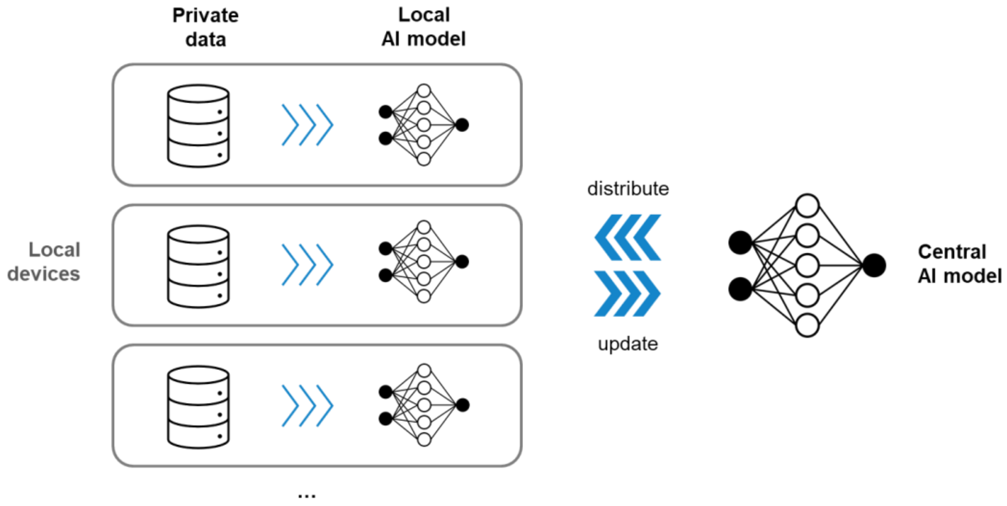

consumers
$f(x)$
Tokenized algorithm
Tokenized results
providers
Data marketplaces based on Smart Contracts
Access control based on Smart Contracts
CtD Environment
owners
Tokenized private data asset
GDPR-Compliance: data asset stays on-premise



Federated Learning on Ocean Protocol

Bringing all of this together, self-sovereign Federated Learning becomes possible on Ocean Protocol. How could a basic FL workflow look like?

1) Data consumers and providers are matched on Federated Learning Portals, like industry specific data markets, that act as aggregators of demand and supply. While data and algorithm providers can publish their offers on data markets, data consumers can visit data markets to discover them and to acquire compute access to tokenized data assets. For example, imagine a healthcare data space, where patients can sell self-sovereign, and privacy-preserving compute access to the private health data on their smartphones. Researchers and scientists could search and filter for suitable data offers (following the same format, data types, and quality, for example, standardized medical records) to be included in their machine learning jobs.
2) The data consumer buys the respective data tokens and buys compute access by spending the tokens.
3) This triggers several Compute-to-Data jobs. The trusted machine learning algorithm is transferred to the edge devices, executed, and only results in the form of machine learning models (or to be more precise: their gradients) are transferred back to the Portal.
4) The global model is aggregated by the Portal or on the premises of the data consumer.



In its simplest implementation, one could iteratively train a global model across many edge devices. Advanced implementations could entail continuous model aggregation and updates, split learning, randomization, gossip, or consensus methodologies.

In any case, the model can be trained on data of multiple data sources,

- Without exposing any private data of any data source
- Without having a central authority that orchestrates the learning process
- While the data owner keeps full control over their data and over algorithms that are trained on it
- While the full process is transparent and written to an immutable audit trail

Applications

While beneficial for a wide range of companies, enterprises and institutions, especially small and medium companies, can benefit significantly from separating the data and the algorithm role. Without the need to build up their own dedicated AI teams, SMEs are enabled to collaborate seamlessly with a multitude of algorithm providers enabling a rich data offering to the mutual benefit.

Individuals are empowered to monetize their data while keeping full ownership and control, minimizing privacy risks.

A glimpse into the future

The most prominent future applications and use-cases of Federated Learning on Ocean Protocol might be in production, advertisement, and healthcare contexts.

Production: Owners of production machines can contribute their sensitive machine, production, and downtime data to form a global predictive maintenance model. The accuracy of such a global model would likely be better than individual models because more data becomes available. With predictive maintenance based on Federated Learning, manufacturers and suppliers can collaboratively optimize downtimes, maintenance windows, and spare parts purchases while keeping their sensitive data private. With Federated Learning on Ocean Protocol, all of this becomes possible soon in decentralized and trustless environments.

Advertisement: Smartphone users store lots of sensitive and private data on their devices. The potential of this data often remains unused because of concerns over security, privacy, and trust. However, the potential is enormous because the more private the data, the better it can be used for targeted advertisement. With Ocean Protocol, users could take the sovereign decision to monetize their private data for the purpose of targeted advertisement. Advertisers could buy compute access via Federated Learning on Ocean Protocol to build better prediction models on sensitive real-world data. With Federated Learning on Ocean Protocol, all of this becomes possible in decentralized and trustless environments, fully self-sovereign.

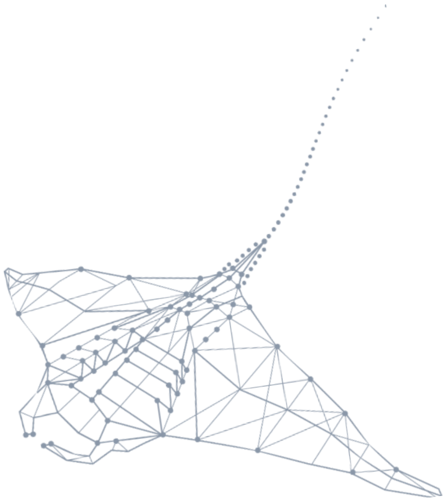
Healthcare: The more data available, the better the prediction model. This is a challenge with use-cases where regulation prevents data sharing, like for example, in the healthcare sector. In the past, medical institutions and researchers had to rely on their own data sources, which can, in addition, be biased by demographics, health insurance, or specialization. Federated learning makes it possible to overcome those challenges by training on multiple distributed data sources, forming a large, distributed training set. Given the properties of Federated Learning on Ocean Protocol, this could be done on a global scale, across continents and cultures, and on edge devices like smartphones, fully controlled by data owners. Nondiscriminatory access to technology and self-sovereign participation can be important first steps towards less biased, more evenly distributed prediction models in healthcare.

Advantages for companies

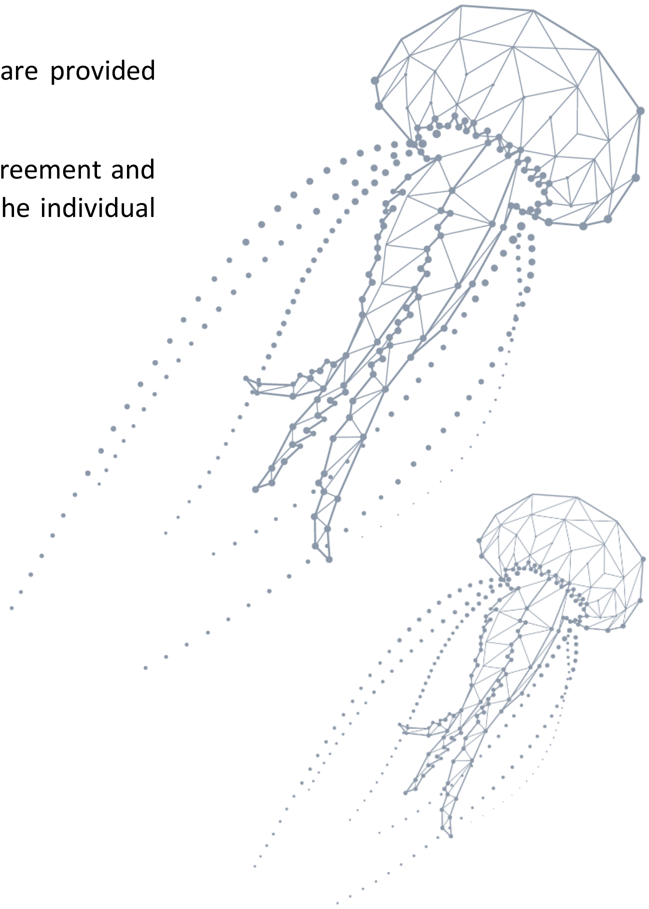
When using Federated Learning on Ocean Protocol, companies and individuals who own data can monetize it without ever disclosing the contents of their data sets. After publishing their data asset on a decentralized marketplace, they can transparently monitor the consumption of their data and keep it in full control, as their data asset stays on-premises.

Federated Learning on Ocean Protocol offers a wide range of advantages for companies. It

1. … helps to significantly reduce the regulatory risk exposure, for example GDPR risks
2. … enables data monetization of very large amounts of data which are often too large to transfer
3. … gives access to new data sources because it empowers anyone to share their sensitive data
4. … enables data monetization without losing control, risk of data loss, copyright infringements, or IP drain as the data never leaves the premise
5. … enables the division of labor between data and algorithm providers and hence facilitates rich data offerings
6. … allows full control over the algorithms that are allowed to access the data



Copyright



Contact

deltaDAO AG
Geibelstraße 46b
22303 Hamburg
Germany

Website	https://delta-dao.com
Mail	contact@delta-dao.com
Twitter	@deltadao
LinkedIn	deltadao
YouTube	deltaDAO