ocean

Ocean Protocol Use Case

Privacy-preserving Data Marketplaces

Privacy-preserving Data Marketplaces based on Ocean Protocol

Abstract

Compute-to-Data on a decentralized marketplace like Ocean Market enables free, open, transparent, secure, and privacy-preserving data sharing and monetization, where data owners keep full control over their data.

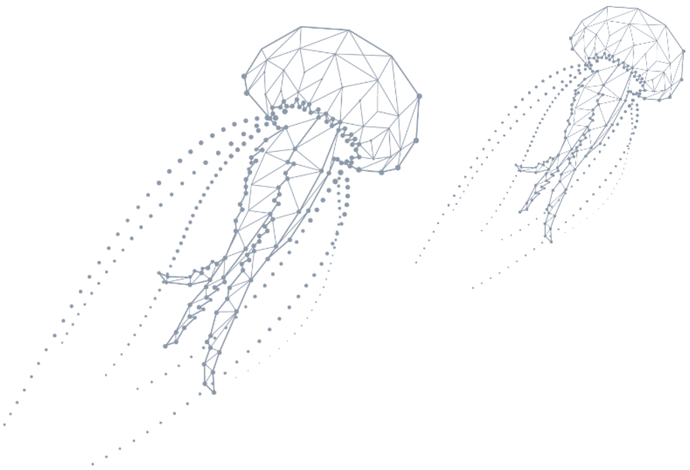
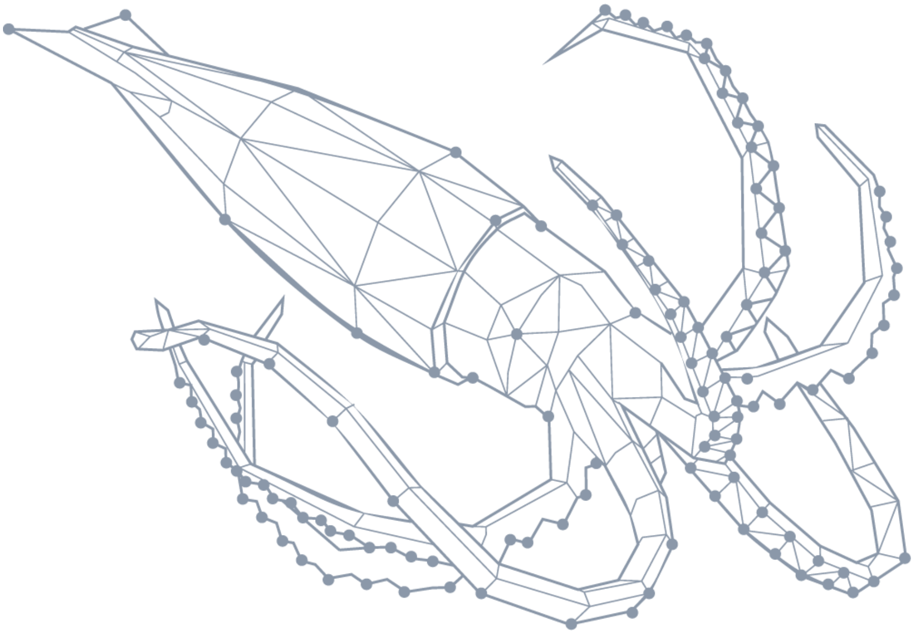


Table of Contents



Introduction

Image: KanawatTH / shuttertstock.com

Tons of valuable data are generated each year. However, substantial amounts of this data remain unexchanged and unused because of concerns about data security and privacy. The lack of trust in central entities controlling the data is another issue, as recent years showed that personal data is intransparently collected and exploited by companies for monetary reasons. Power and control over data shifted from billions of data owners to a few selected companies.

An obstacle to data sharing is particularly present when the data set contains personal data, which at the same time is among the most valuable data. The protection of personal data is essential, of paramount importance, and rightly subject to strict regulatory requirements. Before personal data can be used for monetization or scientific purposes, it must be anonymized. However, anonymization procedures are time-consuming and costly. The lack of published valuable data delays the scientific progress that can result from machine learning.

Centralized data marketplaces evolved, where data owners can share and monetize their data by uploading it to a central authority that operates the marketplace. While organizations and individuals could now decide which data they want to publish and monetize, they again lost control of their data once published on the market. On traditional data marketplaces, data owners still do not know what happens to their data once sold. Moreover, for data sets containing persona data, the anonymization obstacles remain.

Requirements for a privacy-preserving data marketplace

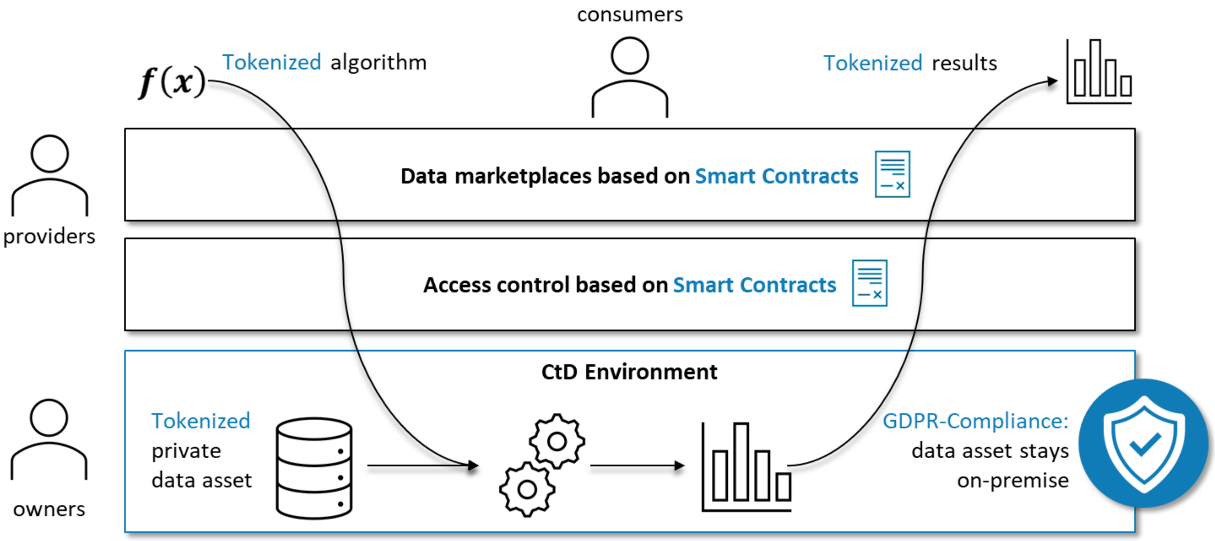
- **Security**: A privacy-preserving data marketplace needs to take measures to secure data in transit and at rest. Access to the data must only be provided to authorized entities.
- **Control**: Moreover, when talking about privacy, user empowerment and data sovereignty are of crucial importance. Data owners should own their own data, decide on their own whether they want to publish it, and should not lose control over their data once they decide to publish it.
- **Transparency**: Data owners should be able to monitor the flow and usage of their data.
- **Anonymization**: Data must be anonymized to protect the privacy of the data subjects.

Ocean Market

Ocean Market is a data marketplace that provides a data access mechanism called Compute-to-data (CtD). CtD enables data owners to grant only compute access to their data. The data itself remains with the data owner in a secured environment.

There are three layers: the data market frontend, the access control layer, and the CtD Environment.

1) Data consumers can visit Ocean Market to search and buy compute access to tokenized data assets. When the data consumer pays the data publisher for the data set with the respective data token, the compute job is triggered.
2) Before the data consumer gets compute access, the access control must be passed. Access control is tokenized and managed by Smart Contracts, stored on-chain.
3) Then, the tokenized and authorized algorithm gets access to the secured environment containing the data asset. The algorithm is executed.
4) After the compute job is done, only the tokenized results are sent back to the asset consumer. The result will no longer contain personal data.


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

Advantages for companies and individuals

Image: *TheDigitalArtist / Pixybay*

When using CtD, 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.

Advantages for publishers and consumers of data assets

CtD is especially suitable to remotely train machine learning models on many data sets owned by different data providers without being dependent on a central authority. To train machine learning models on several personal data sets while preserving privacy benefits many branches, including science and medicine. For instance, research institution A wants to train its AI to detect (or defeat) kinds of diseases. However, A does not own enough valuable data to reach sufficient predictions. The research institutions B-N and the hospitals O-Z, on the other hand, own valuable patient data but do not have AI expertise to implement suitable algorithms. To monetize their data and improve scientific progress, B-N and O-Z provided their data on a decentralized data marketplace. To protect the contents of the data sets, they permit solely compute access of trusted algorithms to their sensitive data. A finds those data sets on the decentralized data marketplace and buys compute access for A's trusted algorithm. A can now train its algorithm on each data set to improve the model and thrive their research. During the entire process of A, no data set was moved, no person working at A ever saw personal data, and no central authority was necessary for orchestration.

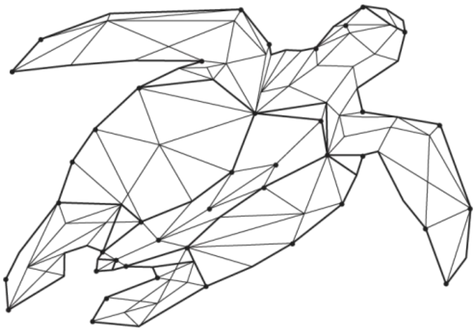
Why CtD on a decentralized marketplace enables a free, open, transparent, secure, and privacy-preserving data sharing and monetization, where data owners keep full control over their data.

	Centralized marketplace without CtD	Centralized marketplace enabling CtD	Decentralized marketplace enabling CtD
(De)centralization	The market is managed by a central entity, which is a valuable target for attacks. Also, centralized marketplaces can be censored.		CtD operates in a trustless and decentralized manner, which obsoletes a central authority for orchestration and offers censorship resistance.
Transparency	It is not transparent to the data owners what happens to their data after they sold it to the marketplace.		DLTs enable auditability, transparency, and trust by design and by default.
Security of data in transit	The data must be secured while transmitting them to the market owner and to the consumers, which poses security risks. The security of the data depends on the market owner.	The data must be secured while transmitting them to the market owner but then will never leave its repository. The security of the data depends on the market owner.	The data never leaves the premises of the data owner.
Security of data at rest	The data is located at the market owner. The security of the data depends on the market owner.		The data is always in possession of the data owner and is there stored in a secured container. The security depends on the data owner and the underlying DLT technology.
Control	Once the data is sent to the data consumer, the consumer can do whatever s/he wants. The control shifted to the data consumer.	The market owner has complete control over their customers, their providers, and the owner's data. The trustfulness of the open algorithms can be approved. The control shifted to the market owner.	The data owner remains the only entity that possesses the data. Data Publishers (owners) can allow selected and trusted algorithms to access their data. So, the data owners never lose control over the data.
Privacy	Personal data must be anonymized before publishing, which is a costly and time-consuming task.	Only aggregated and hence anonymized data is transmitted to the authorized buyers. However, personal data is transmitted to the market owner if the data set was not anonymized.	Only aggregated and hence anonymized data is transmitted to the authorized buyers.

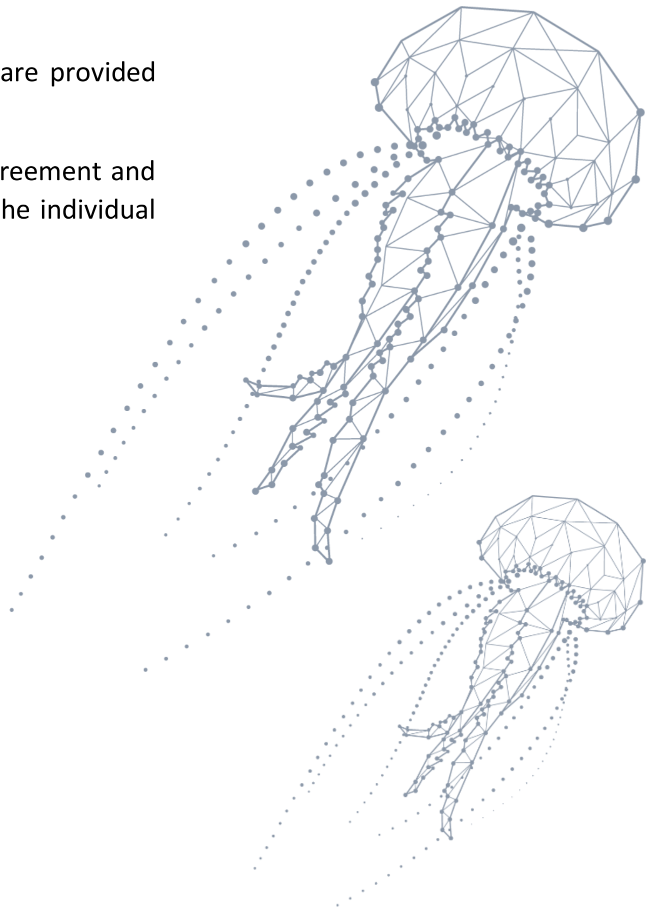
References

Ocean Protocol Foundation Ltd., https://oceanprotocol.com/

Ocean Protocol Developer Documentation, https://docs.oceanprotocol.com/



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