computers

Article

Using Privacy-Preserving Algorithms and Blockchain Tokens to Monetize Industrial Data in Digital Marketplaces [†]

Borja Bordel Sánchez [1,*] , Ramón Alcarria [1] , Latif Ladid [2] and Aurel Machalek [2]

1   Departamento de Sistemas Informáticos, E.T.S.I. de Sistemas Informáticos, Universidad Politécnica de Madrid, C/Alan Turing, s/n, 28031 Madrid, Spain; ramon.alcarria@upm.es
2   Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, 4365 Esch-sur-Alzette, Luxembourg; latif.ladid@uni.lu (L.L.); aurel.machalek@uni.lu (A.M.)
*   Correspondence: borja.bordel@upm.es
†   This paper is a high-quality expansion of the conference article: Sánchez, B.B.; Alcarria, R.; Valladares, T.R. A Data-Driven Blockchain-Based Marketplace to Promote Transparency and Accountability in City-Region Food Systems. In Proceedings of the 2023 18th Iberian Conference on Information Systems and Technologies (CISTI), Aveiro, Portugal, 20–23 June 2023; pp. 1–6.

Abstract: The data economy has arisen in most developed countries. Instruments and tools to extract knowledge and value from large collections of data are now available and enable new industries, business models, and jobs. However, the current data market is asymmetric and prevents companies from competing fairly. On the one hand, only very specialized digital organizations can manage complex data technologies such as Artificial Intelligence and obtain great benefits from third-party data at a very reduced cost. On the other hand, datasets are produced by regular companies as valueless sub-products that assume great costs. These companies have no mechanisms to negotiate a fair distribution of the benefits derived from their industrial data, which are often transferred for free. Therefore, new digital data-driven marketplaces must be enabled to facilitate fair data trading among all industrial agents. In this paper, we propose a blockchain-enabled solution to monetize industrial data. Industries can upload their data to an Inter-Planetary File System (IPFS) using a web interface, where the data are randomized through a privacy-preserving algorithm. In parallel, a blockchain network creates a Non-Fungible Token (NFT) to represent the dataset. So, only the NFT owner can obtain the required seed to derandomize and extract all data from the IPFS. Data trading is then represented by NFT trading and is based on fungible tokens, so it is easier to adapt prices to the real economy. Auctions and purchases are also managed through a common web interface. Experimental validation based on a pilot deployment is conducted. The results show a significant improvement in the data transactions and quality of experience of industrial agents.

Keywords: data-driven marketplace; blockchain; tokenization; monetization; randomization algorithms; transparency

1. Introduction

The future data economy [1] is characterized by extreme exponential business models, where productivity and benefits are no longer related to the amounts of manufactured products, but to the level of knowledge that companies can create or extract [2]. In this context, industrial efficiency is expected to suffer a sharp increase, and costs are to be reduced to the minimum value. New paradigms, such as Industry 5.0 [3], or new economic sectors, such as data services [4], are born and enabled thanks to this data revolution.

However, as in many other previous economic revolutions, the resulting and growing data market is asymmetric [5]. And, against traditional assumptions, some agents have a relevant negotiation position [6]. Thus, many companies cannot compete fairly. On the one hand, data-driven businesses require very specialized digital organizations, where technologies such as Artificial Intelligence or Data Science can be exhaustively exploited.

These technologies allow massive automatization and generate great benefits from third-party data at a very low cost (in fact, the marginal cost is negligible). On the other hand, data are produced by regular companies while they perform common industrial activities. In the traditional economy, these data are valueless sub-products, which are used just as historical records. But their generation is associated with great costs and linked to standard industrial activities. Even when regular industrial companies are aware of the potential value of their data, they have no mechanisms to play an active role in the data market and negotiate a fair distribution of the benefits of these new data-driven businesses [7]. In general, data have no value while they are not processed or integrated in a digital service that is able to extract knowledge from them. Thus, industrial agents are often forced to transfer their data for a reduced price or, more commonly, for free [8].

Although, at first glance, this situation appears to be a market inefficiency which will be adjusted at some point by economic agents, it has a critical impact on the entire society [9]: customers and final users will suffer high prices, which will reduce access to digital services; some industrial sections may disappear; and the strategic independence of countries and continents could be threatened.

Therefore, new data-driven marketplaces are needed to enable fair and active trading between all potential agents in the data economy. These marketplaces should be simple to facilitate the integration of all industrial partners, transparent to promote a fair negotiation, and they should ensure the preservation of industrial secrets (as only authorized buyers should access the data).

In this paper, we propose a blockchain-enabled web-based marketplace to fill this gap. The proposed solution uses the Ethereum network to define a transparent trading scheme controlled by smart contracts. The blockchain network is connected to different services. First, data owners can upload and deposit their data into an Inter-Planetary File System (IPFS) using a web interface. However, before completing the final deposit, the data are randomized to ensure the preservation of industrial privacy. The randomization algorithm is based on numerical stream ciphers, where the number in the stream indicates the random positions of characters in the IPFS. In parallel, the blockchain network receives a transaction and defines a Non-Fungible Token (NFT) to represent the new dataset. Only the NFT owner can request the essential seed to de-randomize the dataset and extract all of the information from the IPFS. Data trading, now transformed into an NFT market, is managed through a web interface where data owners and buyers are active and can define purchases and auctions. All transactions are transparent and public. Additionally, the data market is very volatile. Therefore, transactions in this data-driven marketplace are based on fungible tokens so it is easier to adapt prices to the real economy. Final monetization can be achieved by exchanging cryptocurrencies in any of the existing secondary markets.

The remainder of this paper is organized as follows: Section 2 introduces the state of the art regarding data-driven and blockchain marketplaces. Section 3 describes the proposed marketplace, including the randomization algorithm, the architecture, and the monetization instruments (tokens). Section 4 describes the proposed experimental validation and discusses the results. Section 5 concludes the paper.

2. State of the Art regarding Data-Driven and Blockchain Marketplaces

Blockchain-enabled data marketplaces have been proposed and discussed by many authors in recent years. However, most works propose high-level layered architectures [10] whose real performances are unknown, abstract discussions about the upcoming change in the data management paradigm [11], or theoretical proposals about how communication flows in interactions among participants should be organized [12] in hypothetical future blockchain data marketplaces. Some works even focus on very specific functionalities, such as an optimal pricing algorithm [13], which are described and validated in an isolated environment so that the real performance in a data marketplace is not reported.

However, some small implementations of blockchain data marketplaces may be found. But these solutions are either based on simulations [14] or an initial proof of concept

whose technical implementation is not provided [15], and only the general architecture is described. Some evidence on the performance of these prototypes has been reported, but only emulated scenarios [16] with no real blockchain networks can be found. In some cases, authors have replicated the behavior of blockchain networks using programming interfaces [17], but still, no real information is provided about how these marketplaces would behave in a real ecosystem.

Data marketplaces have been used to enable the monetization of data from Internet-of-Things deployments [18]. In general, these proposals replicate the traditional business model of Internet providers, where users or applications that consume network services must pay a fee [19]. Related challenges such as personal data protection [20] or different monetization models that may increase income [21] have also been studied. But while this scheme could be extended to industrial scenarios, it fits much better applications, such as Smart Cities [22]. In industrial ecosystems, it is difficult for non-technological companies to launch a new division on digital service commercialization. In fact, some authors have explored the challenges faced by data providers through small prototypes and experiments [23]. They concluded that democratization in the access of data markets is still a pending question. Therefore, typical works on business-to-business data transference are oriented to "data sharing" [24] instead of "data monetization", promoting the free use of industrial data by digital companies.

When the blockchain revolution exploded, it was immediately adapted to data marketplaces. In the most common approach, blockchain systems act as brokers for publication/subscription networks [25], where smart contracts apply fees [26]. However, again, the proposed architectures are more suitable for smart environments [27] than for industrial ecosystems [28]. However, as all data must be managed by blockchain nodes, the performance of these marketplaces tends to be poor. Models to facilitate data trading through a blockchain have also been proposed [29], together with rules for coherent and consistent smart contracts [30] and protocols to ensure that delays and quality of service are not affected by heavy consensus algorithms [31]. But the problem is only mitigated, and the final performance is still far from that of other digital platforms [32]. Enhanced versions of these marketplaces, including additional security modules [33] or machine learning [34], have also been reported. However, the structural problems remain unsolved.

As a response, some authors have proposed models where data are not managed by blockchain nodes but stored in distributed databases (such as IPFS) whose consistencies are controlled through smart contracts [35]. The results are promising, although the proposed schemes are fit for very restricted industries, such as city-region food systems. A similar proposal can be found for smart communities [36], irrigation systems [37], electricity grids [38], art products [39], digital libraries [40], or agricultural markets [41]. In addition, proposals to enable users to monetize their personal data have been analyzed [42]. Regardless, more general architecture is needed.

On the other hand, key technologies that allow the implementation of blockchain-enabled marketplaces have also been studied [43]. Contributions have been reported to facilitate the use of specific blockchain solutions, mainly Ethereum [44], and general proposals [45]. Authors have analyzed several different aspects of blockchain networks to enhance their efficiency and promote data trading and monetization. Schemes to reduce the cost of transactions [46] or ease the interaction among industrial agents [12] have been investigated. Technologies that allow for the autonomous operation and interconnection of data marketplaces [47] with cyber-physical systems [48] and other networked devices can also be found [49].

The proposed marketplace in this paper also employs an IPFS to improve the general performance and avoid limitations related to data storage. Additionally, it uses NFTs to generalize the proposal to any kind of data and uses fungible tokens to allow its implementation in any blockchain network. Moreover, the web interface acts as a real marketplace, allowing industrial data owners to register products, prices, auctions, etc.

3. A Blockchain-Based Data-Driven Marketplace

In this section, we present the new data-driven marketplace. In Section 3.1, we introduce the general architecture and global behavior. Section 3.2 discusses the implementation of NFTs and monetization instruments. And finally, Section 3.3 describes privacy-preserving algorithms.

3.1. General Overview

Figure 1 shows the proposed architecture for the blockchain-based data-driven marketplace. Five different segments can be distinguished. The user segment offers a web interface for industrial agents and digital actors to operate the marketplace. It consists of a web server, preferably based on JavaScript technologies, as this language provides stable libraries to interact with a blockchain and distributed storage. Actors within this marketplace can upload new datasets or bid on or buy available data.
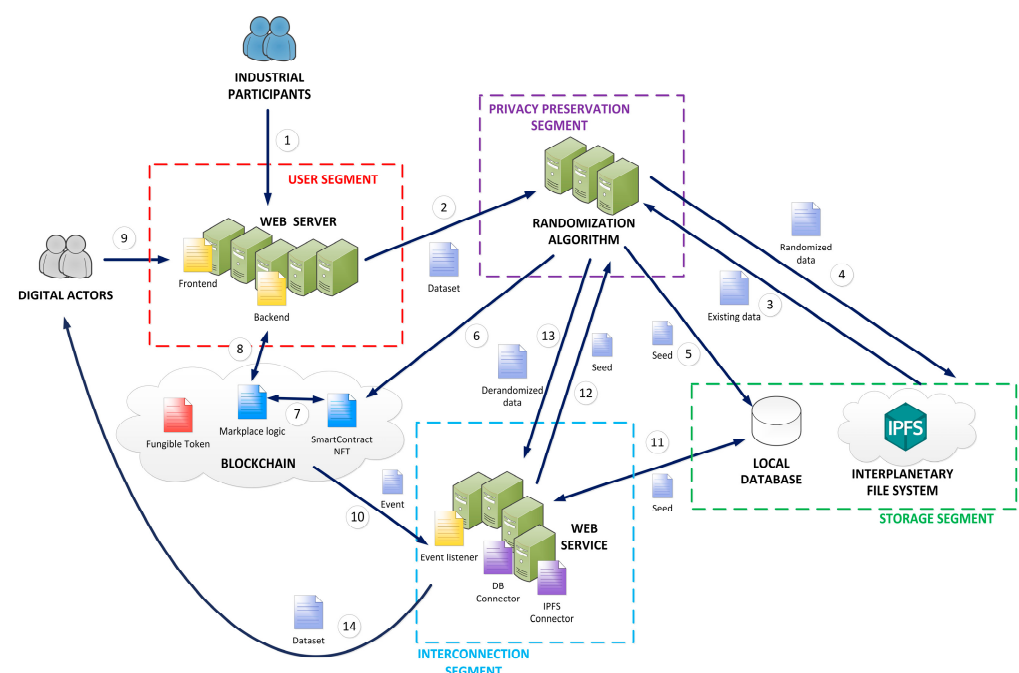

Figure 1. Architecture for data-driven blockchain marketplace.

Industrial partners, as data owners, typically upload data to the marketplace ①. The user segment then has a direct connection to the privacy-preservation segment, which receives the data and applies a randomization algorithm to guarantee industrial secrecy ②. To achieve this, the privacy preservation algorithm extracts previously existing data from the Interplanetary File System (IPFS) in a storage segment ③ and mixes the previous samples with the new ones by following a random variable. The mixed data are stored again in an IPFS ④. The randomization process needs a seed, which acts as a secret key for the security scheme. The seed is stored in a local database within the storage segment ⑤.

Once the randomization and storage processes are completed, a blockchain transaction is triggered by a privacy preservation segment ⑥. This transaction creates an NFT representing the new dataset. This NFT is added to the smart contract where all of the NFTs that are available for purchase are listed ⑦. The blockchain segment is connected to the user segment's backend, so the web interfaces can display all of the available NFTs ⑧.

Industrial partners can decide how their datasets (now NFT) are offered: direct sale, auction, etc. The web interface in the user segment interacts with the blockchain segment to adapt the internal logic of every NFT to the owners' preferences ⑧.

On the other hand, the digital actor can use the same web interface to review the offered datasets and acquire them ⑨. The transaction is supported by digital wallets and fungible

tokens (see Section 3.2). Fungible tokens can be acquired in a secondary market depending on the monetization model proposed for each scenario or marketplace implementation.

Once the NFT owner is changed in the blockchain system, the new proprietary can use the blockchain function to extract and obtain the raw data. The invoked transaction creates an event, which is received by the interconnection segment ⑩. The interconnection segment makes the event-oriented architecture of the blockchain networks and the remote call architecture in common services and servers compatible. It is typically a web service. It transforms the blockchain event into a call to recover the proper seed from the storage segment ⑪. This seed is forwarded to the privacy-preserving algorithm ⑫, which, with that information, derandomizes the information from the IPFS module ⑬. The data are returned to the NFT owner directly from the interconnection segment ⑭ to avoid heavy data transferences through the blockchain.

This workflow is repeated for each dataset to be monetized in the marketplace.

The user interface will be public and will include secondary functions, such as registering new participants or visualizing a historical register of transactions. The backend offers a REST API (Representational State Transfer Application Programming Interface), so the marketplace can be integrated into other platforms such as mobile devices. Figure 2 shows the proposed interface.
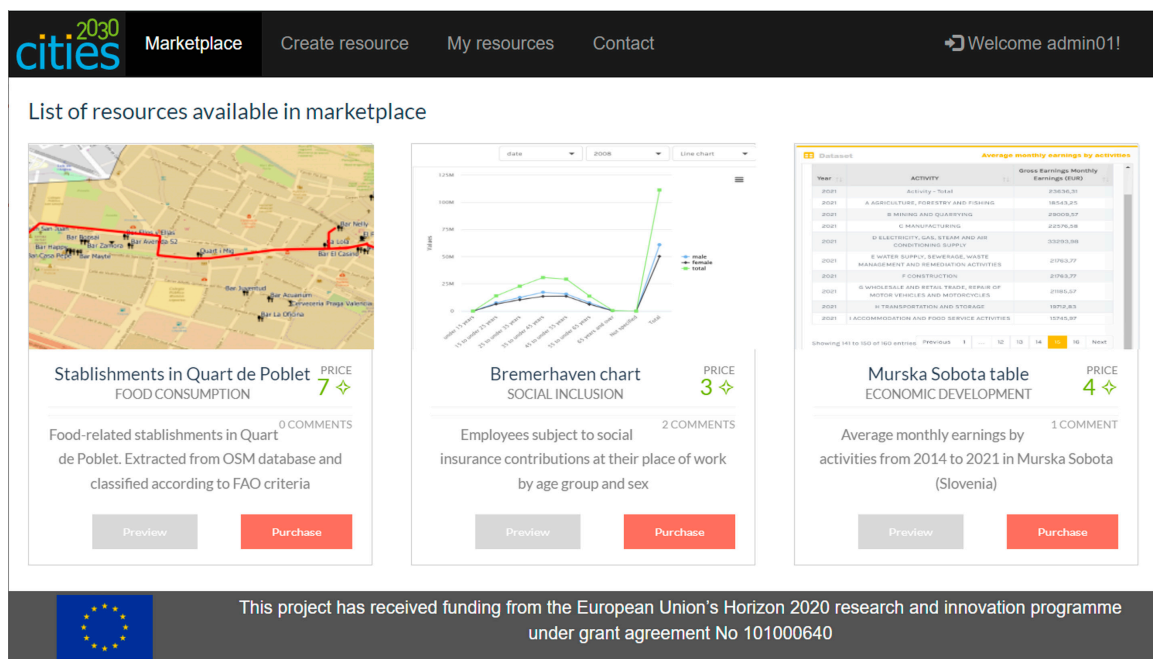

Figure 2. The web interface in the user segment.

As a web module, it can create automatic connections to several digital wallets, so industrial and digital agents can operate the marketplace and their NFTs and fungible tokens in a standardized way. Digital wallets are installed as extensions in web browsers, reducing the complexity of the user segment. This offers additional advantages, such as the ability to use several crypto accounts or shared devices.

Regarding the blockchain network, existing public systems can be used. But for most solutions, the code is open, and the deployment works are not complex. Thus, a private blockchain network can be used as well if this business model is considered more suitable for some application scenarios. However, several authors have warned about the risks associated with these private networks [50], as they need significant effort in software maintenance and operation. This network uses smart contracts (or other similar programming instruments) to define fungible tokens that are used as cryptocurrency in the marketplace (see Section 3.2). Fungible tokens are implemented and initialized using

a common function and invocation. As a result, a certain number of interchangeable and indistinguishable tokens are created. On the other hand, an NFT has a unique identifier, and the value can change from one token to another.

The interconnection segment is a collection of services, where the main one is an event listener. This listener is waiting for the event from the blockchain network. For each event, it establishes a connection to the storage segment through the proper connector. In general, there is a different connector for every storage module and technology in the storage segment.

In the storage segments, datasets and all persistent data structures are maintained. In general terms, this segment includes two modules. First, a local database is considered, where all of the seeds employed to randomize each of the datasets in the marketplace are considered. The database can be relational or non-relational. Second, IPFS technology is used for the preservation of datasets. An IPFS is a platform and a protocol used to store and share hypermedia through a distributed file system. It is specially designed to run together with distributed applications, such as the proposed marketplace. Somehow, an IPFS is similar to the World Wide Web, where a distributed hash table and a centralized name space guarantee that all content is accessible and reachable.

3.2. Token and Monetization Instruments

Data monetization in the proposed architecture is based on fungible tokens. Fungible tokens are interchangeable, identical software objects which can be used to represent a proprietary cryptocurrency. Although there are several different approaches to define and create a fungible token, in our marketplace, we follow the ERC20 standard [51]. This standard is usually considered the most suitable scheme to define a fungible token. As they are non-unique, a large number of fungible tokens are generated through the same invocation. All of them have identical values. They are therefore the perfect instruments to define cryptocurrency.

The ERC20 standard was originally proposed in 2015. This standard is designed to operate in Ethereum networks, although it can be adapted to other blockchain technologies. The standard standardizes the attributes and functions to be implemented in a fungible token to ensure their operability. The ERC20 standard considers eight mandatory functions and three additional optional ones. In the proposed architecture, fungible tokens are programmed by encapsulating the following functions (see Figure 3):

```solidity
1   pragma solidity ^0.8.20;
2   interface IERC20 {
3       function totalSupply() external view returns (uint);
4       function balanceOf(address account) external view returns (uint);
5       function transfer(address recipient, uint amount) external returns (bool);
6       function allowance(address owner, address spender) external view returns (uint);
7       function approve(address spender, uint amount) external returns (bool);
8       function transferFrom(address sender,address recipient,
9           uint amount) external returns (bool);
10      event Transfer(address indexed from, address indexed to, uint value);
11      event Approval(address indexed owner, address indexed spender, uint value);
12  }
```

Figure 3. Token implementation using the ERC20 standard.

- *name*: This returns the name of the token. Although this function is optional, we decided to include it to improve usability. However, in other marketplace implementations, interfaces and smart contracts must not expect the function to be present under all circumstances. In this first marketplace implementation, the token was named "Cities2030 token".
- *symbol*: This returns the symbol of the token. As in the previous one, this is an optional function. But we included it for clarity purposes. In this first data-driven marketplace implementation, the symbol for the fungible token was ✧.

- *decimals*: This is also an optional function and returns the number of decimals that the token uses. As the proposed solution has monetization purposes, the use of a decimal facilitates economic transactions and negotiation.
- *totalSupply*: This returns the total token supply.
- *balanceOf*: This is employed to obtain the account balance of another user, so it can be decided whether the transaction is feasible or not.
- *transfer*: This transfers a given number of tokens to the indicated user (address) and triggers an event to notify the transaction.
- *transferFrom*: This transfers the indicated number of tokens from a referred original account (or address) to the target address (account). It also triggers an event to notify the transaction.
- *approve*: This allows a user to withdraw from the account multiple times up to the maximum amount indicated in the function.
- *allowance*: This returns the amount that a user is still allowed to withdraw from the target account as a charge for a bought dataset.

The ERC20 standard, on the other hand, employs two different types of events, which can be used to define the back-end logic of the web interface and the interconnection segment. The two events are as follows:

- *Transfer*: This must be triggered when tokens are transferred, including zero-value transfers.
- *Approval*: This must be triggered on any successful call to the *approve* function.

On the other hand, the proposed marketplace employs an NFT as well to represent datasets and operate with their ownership within the blockchain network. In this case, we also propose the use of a standard: the ERC721 standard [52].

The key attribute for any NFT created according to the ERC721 standard is the *tokenId* variable. This identifier must be unique in the entire system and allows any smart contract or software piece to refer to and operate with this NFT. The functions (optional and mandatory) in both standards (ERC20 and ERC721) are pretty similar, apart from the unique identifier and the "ownership" property. Figure 4 represents an NFT implementation using the ERC721 standard. In ERC721 tokens, as they are unique, ownership can be defined. We take advantage of this property to manage seeds in the privacy preservation segment. The function ownerOf is used to identify the current owner of the NFT, labeled *tokenId*.

```solidity
1    pragma solidity ^0.8.20;
2    interface IERC721 {
3        function balanceOf(address owner) external view returns (uint balance);
4        function ownerOf(uint tokenId) external view returns (address owner);
5        function safeTransferFrom(address from, address to, uint tokenId) external;
6        function safeTransferFrom(address from,address to,uint tokenId,
7            bytes calldata data) external;
8        function transferFrom(address from, address to, uint tokenId) external;
9        function approve(address to, uint tokenId) external;
10       function getApproved(uint tokenId) external view returns (address operator);
11       function setApprovalForAll(address operator, bool _approved) external;
12       function isApprovedForAll(address owner,
13       address operator) external view returns (bool);
14   }
```

Figure 4. NFT implementation using ERC721 standard.

But an NFT typically must implement other functionalities in addition to the basic ones. Specifically, it must implement the functions inherited from the application scenario and use the cases that will be covered by the marketplace. In order to combine standard functions and domain-specific functions, the NFT must also comply with the ERC165 standard [53].

The ERC165 standard is a mechanism to automatically detect and publish the interface that any smart contract implements. A JSON-based description is generated and exchanged

when ERC165 functions are invoked. For example, in the Ethereum context, this description is usually known as the ABI (Application Binary Interface). When any smart contract is a direct implementation of a public standard, the JSON description just indicates the version. In contrast, the full interface description must be shared among all interconnected smart contracts. The ERC165 standard includes only one function, indicating whether or not the NFT implements (boolean output) a given ABI. In the standard, this function is named *supportInterface*.

Given that the ERC20 standard and the ERC721 standard share several common functions, many codes could be reused, reducing the system complexity and fees associated with contract deployment. The ERC1155 standard [54], known as the "multitoken standard", is used for this purpose. This standard allows for the definition of a fungible token and NFT using the same smart contract. In the proposed marketplace, we use this approach because of its efficiency. Figure 5 shows the structure of the ERC1155 multitoken standard in a flowchart.
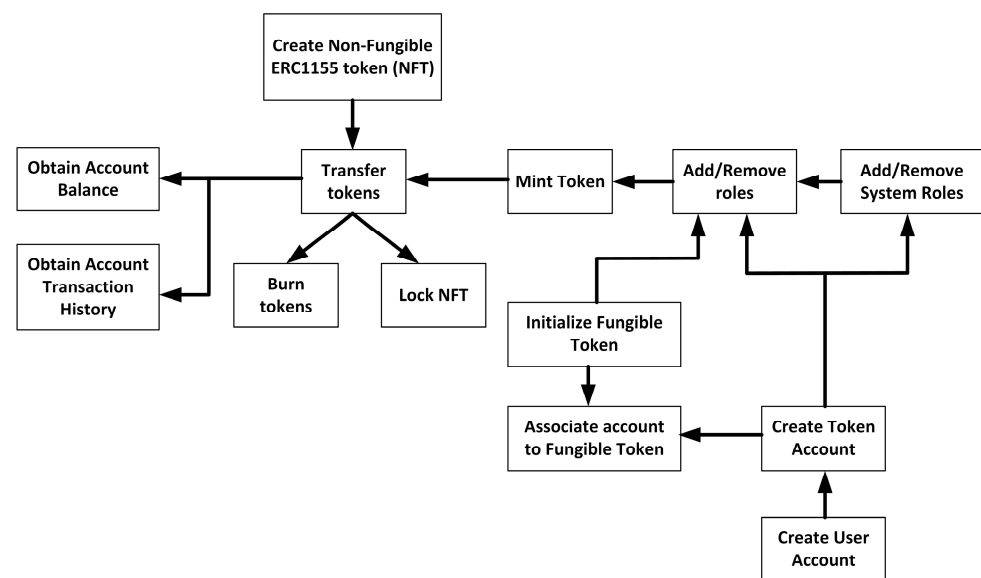

Figure 5. ERC1155 standard flowchart.

3.3. Privacy-Preserving Algorithm

The proposed privacy preservation strategy consists of a random mixing scheme. Information units (letter, sentences, data samples, etc.) are randomly reordered, so recovering the dataset in the original form without the proper key is not feasible. In order to increase the level of randomization and privacy, fictitious information is also injected from previously stored datasets in the IPFS modules.

A dataset D is composed of N different information units. This dataset is transformed into a new randomized dataset D_{rand} including N_{rand} information units, where $N_{rand} > N$. Additionally, we assume that a collection C of M information units are extracted from the IPFS module to enrich the randomization process.

Function $q = \{(p)$ maps data samples in dataset D into samples in dataset D_{rand}, so a sample in the p-th position in dataset D is moved to the q-th position in dataset D_{rand}. This function is a Pseudorandom Number Generator (PRNG). Because it is computationally

lightweight and easy to implement, in the proposed privacy-preserving algorithm, we use the Trifork PRNG (1) [55]. Figure 6 presents the architecture of this PRNG.

$$
\begin{aligned}
q &= \{(p) = w_p \\
x_p &= \left((x_{p-r1} + x_{p-s1}) \bmod \alpha \right) \oplus z_p^{\gg} \\
y_p &= \left((y_{p-r2} + y_{p-s2}) \bmod \alpha \right) \oplus x_p^{\gg} \\
z_n &= \left((z_{p-r3} + z_{p-s3}) \bmod \alpha \right) \oplus y_p^{\gg} \\
x_p^{\gg} &= \left((x_{p-r1} + x_{p-s1}) \bmod \alpha \right) \gg \beta \\
y_p^{\gg} &= \left((y_{p-r2} + y_{p-s2}) \bmod \alpha \right) \gg \beta \\
z_p^{\gg} &= \left((z_{p-r3} + z_{p-s3}) \bmod \alpha \right) \gg \beta \\
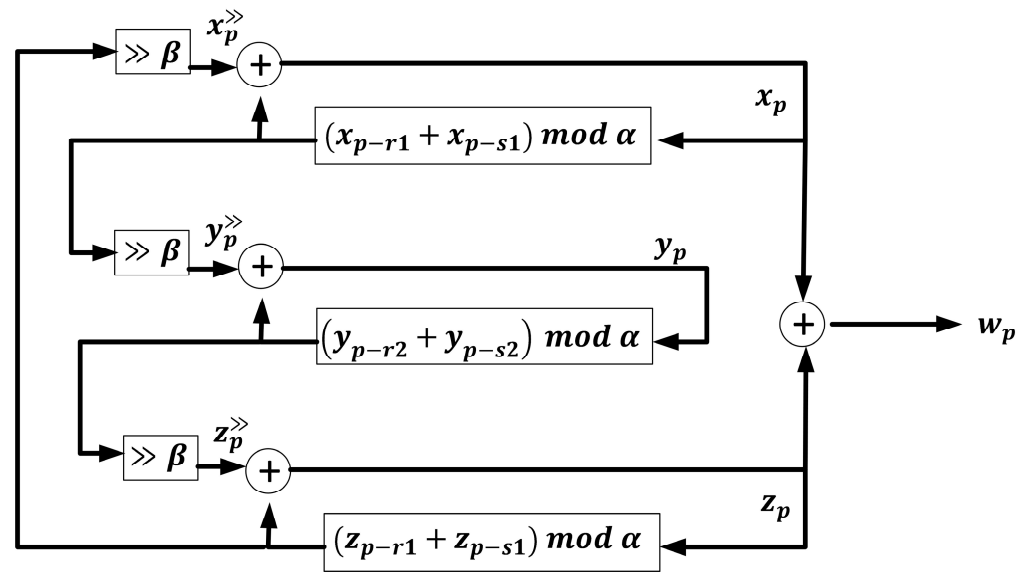w_p &= x_p \oplus z_p
\end{aligned}
\tag{1}
$$

Figure 6. Trifork architecture.

Variables α, β, $r_{\{1,2,3\}}$, and $s_{\{1,2,3\}}$ are configuration parameters. First, α must be chosen to ensure that the Trifork PRNG is able to produce enough numbers to map all of the N_{rand} positions in the randomized dataset D_{rand} (2). Later, as the right-shift and XOR operations are bit-oriented, the β parameter must be carefully chosen so that the value is lower than the maximum number of bits in random numbers w_p (3). Finally, the $r_{\{1,2,3\}}$ and $s_{\{1,2,3\}}$ values can be freely selected.

The seed is composed of three vectors of information units, x_{init}, y_{init}, and z_{init}, whose lengths depend on the selected values for $r_{\{1,2,3\}}$ and $s_{\{1,2,3\}}$ (4).

$$
\alpha \geq N_{rand} \tag{2}
$$

$$
\beta \leq \lfloor \log_2 \alpha \rfloor \tag{3}
$$

$$
\begin{aligned}
length\{x_{init}\} &= max\{r_1, s_1\} \\
length\{y_{init}\} &= max\{r_2, s_2\} \\
length\{z_{init}\} &= max\{r_3, s_3\}
\end{aligned}
\tag{4}
$$

Other PNRGs could be used to support this privacy-preserving scheme. However, Trifork offers some essential advantages over other existing generators in the state of the art. First, as datasets may be large, it is critical that the selected PRNG can produce very long random streams where no number or sequence is repeated (as that would cause a conflict when accessing the file system). Trifork, compared to other reported PRNGs, has a very large key space. Some studies [56] have shown that Trifork multiplies the length of random streams several times compared to other common technologies. Second, the processing

time should not affect the quality of experience and must be negligible compared to the standard data transmission delays. Trifork is known for its very reduced computational cost and scalability compared to any other PRNG previously reported [57]. As it is only based on simple binary operations, the processing delay is almost negligible. So, it is the most suitable technology for our marketplace.

After all, N original information units are placed in the randomized dataset D_{rand}, and the remaining $N_{rand} - N$ positions are filled with random information units from the collection C extracted from the IPFS module. This collection can be generated through a second PRNG or any other instrument for random access. When the randomized dataset is ready, it is stored in an IPFS again. And the seed is saved in the local database in the storage segment.

This privacy-preserving and data randomization algorithm ensures data security against unauthorized accesses or attacks against data storage. Even if authorized access is granted, data cannot be derandomized without the proper seed (key) and number stream. Although a brute force attack is still possible, the elevated entropy of the Trifork PRNG guarantees that the probability of making this attack successful is negligible. Moreover, the blockchain network and solidity language ensure that only the NFT owner obtains the seed to derandomize the data, as function modifiers to control access are embedded into the blockchain architecture and are now considered fully secure.

4. Experimental Validation and Results

In order to evaluate the performance of the proposed marketplace, an experimental validation process is designed and carried out. The experimental validation process was based on pilot experience and deployment. The proposed marketplace was implemented in the context of the Arganda living lab within the Cities2030 project.

The Cities2030 project is funded by the European Commission with the objective of defining new good practices, policies, and technologies to improve city-region food systems. To achieve this, fifteen different living labs were established in more than 12 European countries. The pilot marketplace was implemented in the Arganda lab (Spain). This living lab aims to explore new digital instruments to improve efficiency and monetization in the food processing industry. Previous experiments on controlling work conditions [28] or increasing transparency in the food supply chain [58] were successful in this context.

For this pilot, the user segment in the marketplace was implemented using Node.js technology (JavaScript). To be more specific, the Express framework was employed to create the entire segment. Node.js is a light technology that shows great scalability, so it is perfect for this kind of distributed application.

The blockchain network was based on Ethereum technology. But for this prototype, we do not use the public Ethereum network nor a private full implementation. Instead of that, we use Ganache as an Ethereum network. Ganache is a software suite for local and personal blockchain application development over Ethereum and Corda. It includes a graphic interface and a command line interface. Ganache offers the same API as the Ethereum network, but no distributed nodes must be deployed, and only a local process handles all transactions. This facilitates the creation of small pilots and prototypes.

The interconnection segment and the privacy preservation segment were implemented using Node.js, too. Libraries such as web3 allow for the integration of external services (event listeners, database managers, etc.) with blockchain in JavaScript environments. Then, the MongoDB database was used for local storage, as it can be easily operated through JavaScript scripts. MongoDB is a noSQL database oriented to collections and JSON objects.

The marketplace was deployed in a large industrial company belonging to the bakery sector in the context of the Cities2030 Arganda living lab. The company had more than three hundred workers who were free to use the marketplace according to their needs for three months. No control was proposed or established, and all participants could use the marketplace without supervision. A user manual and a video recording with guidelines

were included to facilitate the use of this new tool. The research group was only available for incidents.

Participants in the pilot experience were divided into two groups with similar population characteristics. The first group was employed as the control group to ensure the significance of the reported improvements. The second group was employed as the pilot group. The experience was three months long. After the experience, both groups were asked to respond to a short survey with nine different questions to be answered using the Likert scale (with answers ranging from "Totally agree" to "Totally disagree"). The responses were collected and processed using the MATLAB 2022a software. The Mann–Whitney U test was used to detect relevant differences between the control and pilot groups.

During the experience, statistics about the use of the marketplace were collected to clarify the level of experience achieved by the participants. In addition, the processing delay required to upload or download a dataset was also monitored. Logs from the blockchain network, the web interface, and the IPFS storage were used for this purpose. As delays are proportional to the size of the datasets, the results were normalized to make all transactions comparable.

Table 1 shows the usage statistics that were collected after the three-month experience.

Table 1. Usage statistics.

Parameter	Value
Uploaded datasets (NFT)	157
Registered users	37
Transactions	5495
Total budget (approved operations)	34,069 ✧

Where ✧ represents the ERC20 fungible token defined for the system.

As can be seen, the usage level is relevant. Almost two new datasets were uploaded to the marketplace every year (1.8 datasets, to be precise). And more than sixty (60) transactions were also carried out every day. These results are comparable to the statistics reported in other similar experiences [35], so we can conclude that the usage level is relevant. In addition, before the implementation of the marketplace, companies within the Arganda living lab were unable to monetize their internal datasets. The reported results represent a significant improvement in the number of data transactions and monetization in industrial companies.

Figure 7 shows a histogram where the processing delay per information byte in the proposed marketplace is represented. As can be seen, the most common processing delay is four milliseconds per uploaded byte. The maximum observed delay is around the double value (eight milliseconds per uploaded byte). It can also be seen that the aggregated probability concentrates a little bit around higher delay values, but it has a reduced effect. Compared to other local storage solutions (such as SQL databases), the delays are slightly higher. But this is common in IPFSs and blockchain networks. So, in general, the proposed marketplace has an average performance.
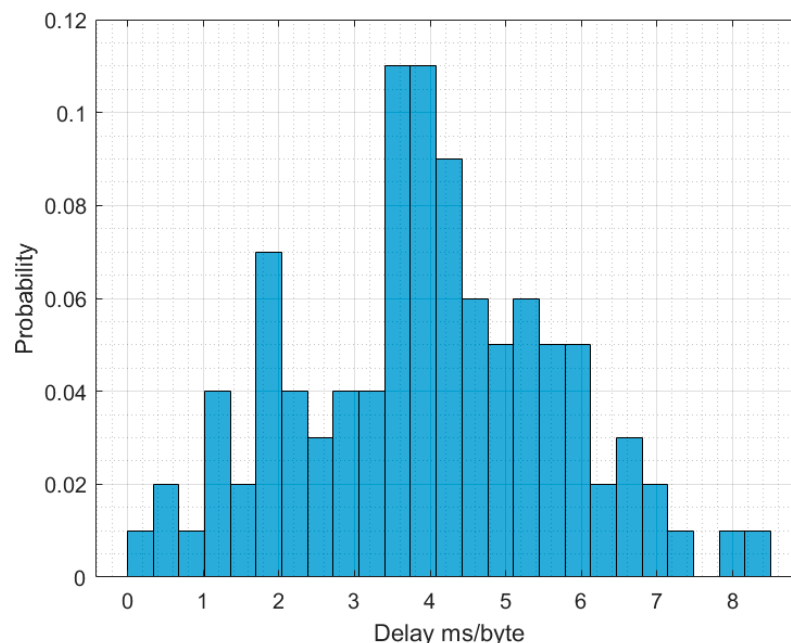
Figure 7. Processing delay (distribution) results.

On the other hand, to analyze the improvement in the quality of experience of industrial agents, we used the Mann–Whitney U test and the responses from the surveys. Different significance thresholds were considered. Table 2 shows the results obtained.

Table 2. Survey results.

Question ID	Question	Significance
Q1	I'm not worried about industrial secrets and conflict of interest	**
Q2	I'm willing to share my data if they are monetized	**
Q3	I'm prepared to upload my datasets to the marketplace	***
Q4	I appreciate transparent and accountable data from industrial partners	**
Q5	I feel encouraged to participate in the data economy	*
Q6	I feel the marketplace is accountable	***
Q7	I would promote the use of this marketplace among other industrial partners	**
Q8	I feel my data are properly monetized	**
Q9	Globally, the quality of my experience with data sharing has improved	**

NS, not significant; * significant at $p < 0.05$; ** significant at $p < 0.005$; *** significant at $p < 0.001$.

As can be seen, there is a significant improvement in all of the questions and in the global quality of experience. Questions related to the accountability and trustworthiness of the market are those where the biggest improvement is detected. The transparency and immutability of blockchain technology is clearly well received by industrial actors. Although most participants surely did not have to check the immutability or transparency of transactions in the marketplace, the experience was improved by the psychological confidence and expectations of users about blockchain networks. On the other hand, although this was significant as well, the participation of industrial companies in the data economy had a less significant improvement. The quality of experience is a subjective indicator, so it is affected by personal thoughts and external factors. For example, in this

case, barriers to participate in this new economic paradigm may not be just technological, but also financial, legislative, etc., so the impact caused by the marketplace is more modest. Although it is an enabling tool, industrial agents still have other challenges in mind which cannot be addressed by the proposed marketplace but affect the global quality of experience.

All of the remaining six questions show an improvement with an equivalent significance level. Figure 8 shows the results from the pilot and control groups in a boxplot in order to allow a deeper discussion.
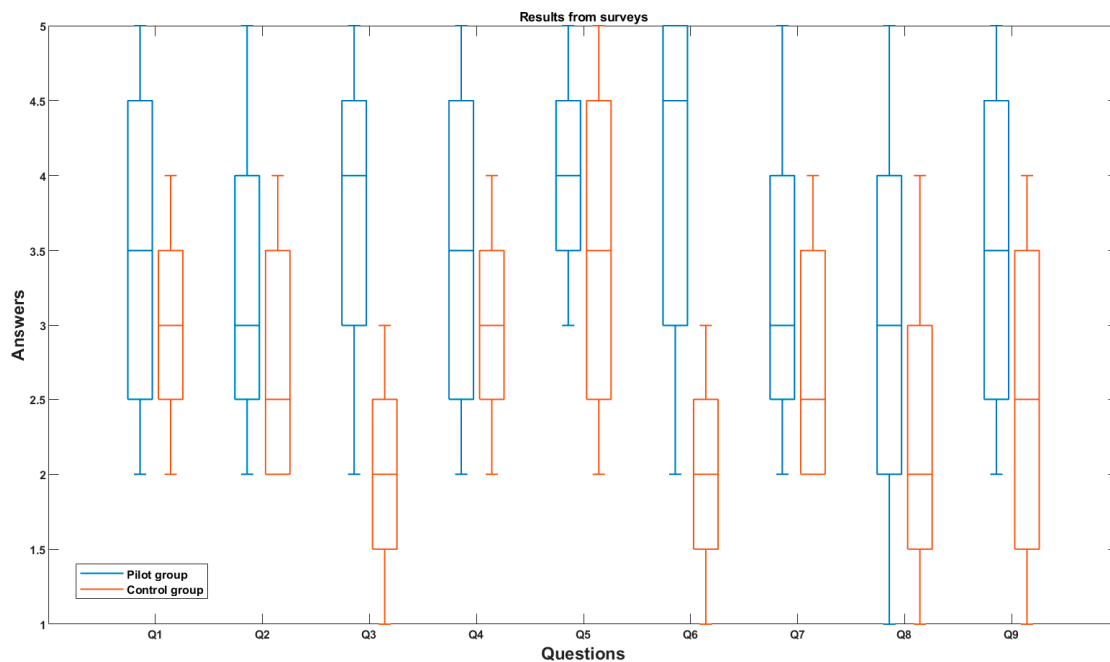

Figure 8. Distribution of answers in survey (results).

Specifically, industrial companies report an improvement in data monetization (which is now perceived as fair). This is a critical point, as it is one of the main objectives of the proposed marketplace. As data owners can negotiate the value of their own data, the satisfaction is surely greater. Figure 6 shows that the average response is more positive in the pilot group (for all questions) than in the control group. This represents a clear improvement in the general quality of experience. If we focus on Q8 (monetization), we can see that the dispersion is high in both the control group and the pilot group. However, only in the pilot group, some individuals' answers were "Totally agree".

In addition, the marketplace has created a solid reputation, and companies are confident that they will recommend other industrial agents to join this new digital platform. To promote the future adoption of the data-driven marketplace, interpersonal communication is essential. So, this is the first evidence of the potential success of this new monetization scheme. Actually, as seen in Figure 8, the results for Q7 show that only the responders in the pilot group provided a "Totally agree" answer. In addition, the average answer is a half-point more positive for the pilot group.

Focusing on the promotion of an innovative ecosystem for open data, it can be seen that Q2 in Figure 8 shows that participants in the pilot group feel more ready to share their industrial data. In this case, the impact of the proposed marketplace as a trustworthy instrument for sharing data according to the preferences of data owners is clear. The same conclusions may be extracted if we analyze the results for Q1. While the participants in the control group show a higher level of worry about the use of their industrial data and the possible leak of industrial secrets, these concerns are not as relevant in the pilot group, where some participants even gave a "Totally agree" answer.

Therefore, we can conclude that the proposed data-driven marketplace is technically sound and promotes an improvement in the quality of experience of industrial companies.

5. Conclusions

In this paper, we propose a blockchain-enabled solution to monetize industrial data. Industries can upload their data to an Inter-Planetary File System (IPFS) using a web interface, where the data are randomized through a privacy-preserving algorithm. In parallel, the blockchain network creates a Non-Fungible Token (NFT) to represent the dataset. So, only the NFT owner can obtain the required seed to derandomize and extract all data from the IPFS. Data trading is then represented by NFT trading and is based on fungible tokens, so it is easier to adapt prices to the real economy. Auctions and purchases are also managed through a common web interface.

Experimental validation based on a pilot deployment was conducted. The results show a significant improvement in the quality of experience of industrial agents and a relevant increase in the number of transactions related to the data economy.

In future work, larger prototypes and experiences will be developed, involving a higher number of participants from different cultures and from the international context.

Author Contributions: Conceptualization, R.A.; methodology, B.B.S.; software, R.A.; validation, B.B.S.; formal analysis, B.B.S.; investigation, B.B.S. and R.A.; resources, L.L.; data curation, L.L. and A.M.; writing—original draft preparation, B.B.S.; writing—review and editing, L.L. and A.M.; visualization, A.M.; supervision, B.B.S.; project administration, L.L. and A.M.; funding acquisition, R.A. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The datasets presented in this article are not readily available due to privacy protection.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Cong, L.W.; Xie, D.; Zhang, L. Knowledge accumulation, privacy, and growth in a data economy. *Manag. Sci.* **2021**, *67*, 6480–6492. [CrossRef]
2. Lammi, M.; Pantzar, M. The data economy: How technological change has altered the role of the citizen-consumer. *Technol. Soc.* **2019**, *59*, 101157. [CrossRef]
3. Bordel, B.; Alcarria, R.; de la Cal Hacar, G.; Valladares, T.R. Efficient and Accountable Industry 5.0 Production Scheduling Mechanism for Mass Customization Scenarios. In Proceedings of the International Conference on Ubiquitous Computing and Ambient Intelligence, Riviera Maya, Mexico, 28–29 November 2023; Springer Nature: Cham, Switzerland, 2023; pp. 44–56.
4. McConaghy, T. Ocean Protocol: Tools for the Web3 Data Economy. In *Handbook on Blockchain*; Springer International Publishing: Cham, Switzerland, 2022; pp. 505–539.
5. Ross, P.; Maynard, K. Towards a 4th industrial revolution. *Intell. Build. Int.* **2021**, *13*, 159–161. [CrossRef]
6. Koutroumpis, P.; Leiponen, A.; Thomas, L.D. Markets for data. *Ind. Corp. Chang.* **2020**, *29*, 645–660. [CrossRef]
7. Steinberg, E. Big data and personalized pricing. *Bus. Ethics Q.* **2020**, *30*, 97–117. [CrossRef]
8. Silva, A.J.; Cortez, P.; Pereira, C.; Pilastri, A. Business analytics in Industry 4.0: A systematic review. *Expert Syst.* **2021**, *38*, e12741. [CrossRef]
9. Hummel, P.; Braun, M.; Tretter, M.; Dabrock, P. Data sovereignty: A review. *Big Data Soc.* **2021**, *8*, 2053951720982012. [CrossRef]
10. Gupta, P.; Dedeoglu, V.; Kanhere, S.S.; Jurdak, R. Towards a blockchain powered IoT data marketplace. In Proceedings of the 2021 International Conference on COMmunication Systems & NETworkS (COMSNETS), Bangalore, India, 5–9 January 2021; pp. 366–368.
11. Maher, M.A.; Khan, I.A. From sharing to selling: Challenges and opportunities of establishing a digital health data marketplace using blockchain technologies. *Blockchain Healthc. Today* **2022**, *5*, 1–9. [CrossRef]
12. Yoo, H.; Ko, N. Blockchain based data marketplace system. In Proceedings of the 2020 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Republic of Korea, 21–23 October 2020; pp. 1255–1257.
13. Liu, K.; Qiu, X.; Chen, W.; Chen, X.; Zheng, Z. Optimal pricing mechanism for data market in blockchain-enhanced Internet of Things. *IEEE Internet Things J.* **2019**, *6*, 9748–9761. [CrossRef]
14. Nasonov, D.; Visheratin, A.A.; Boukhanovsky, A. Blockchain-based transaction integrity in distributed big data marketplace. In Proceedings of the Computational Science–ICCS 2018: 18th International Conference, Wuxi, China, 11–13 June 2018; Springer International Publishing: Cham, Switzerland, 2018; pp. 569–577.

15. Klaine, P.V.; Xu, H.; Zhang, L.; Imran, M.; Zhu, Z. A privacy-preserving blockchain platform for a data marketplace. *Distrib. Ledger Technol. Res. Pract.* **2023**, *2*, 1–16. [CrossRef]

16. González, V.; Sánchez, L.; Lanza, J.; Santana, J.R.; Sotres, P.; García, A.E. On the use of Blockchain to enable a highly scalable Internet of Things Data Marketplace. *Internet Things* **2023**, *22*, 100722. [CrossRef]

17. Sahoo, S.; Halder, R. Traceability and ownership claim of data on big data marketplace using blockchain technology. *J. Inf. Telecommun.* **2021**, *5*, 35–61. [CrossRef]

18. Mišura, K.; Žagar, M. Data marketplace for Internet of Things. In Proceedings of the 2016 International Conference on Smart Systems and Technologies (SST), Osijek, Croatia, 12–14 October 2016; pp. 255–260.

19. Özyilmaz, K.R.; Doğan, M.; Yurdakul, A. IDMoB: IoT data marketplace on blockchain. In Proceedings of the 2018 Crypto Valley Conference on Blockchain Technology (CVCBT), Zug, Switzerland, 20–22 June 2018; pp. 11–19.

20. Koutsos, V.; Papadopoulos, D.; Chatzopoulos, D.; Tarkoma, S.; Hui, P. Agora: A privacy-aware data marketplace. *IEEE Trans. Dependable Secur. Comput.* **2021**, *19*, 3728–3740. [CrossRef]

21. Azcoitia, S.A.; Laoutaris, N. A survey of data marketplaces and their business models. *ACM SIGMOD Rec.* **2022**, *51*, 18–29. [CrossRef]

22. Ramachandran, G.S.; Radhakrishnan, R.; Krishnamachari, B. Towards a decentralized data marketplace for smart cities. In Proceedings of the 2018 IEEE International Smart Cities Conference (ISC2), Kansas City, MO, USA, 16–19 September 2018; pp. 1–8.

23. Eichler, R.; Gröger, C.; Hoos, E.; Schwarz, H.; Mitschang, B. From data asset to data product–the role of the data provider in the enterprise data marketplace. In Proceedings of the Symposium and Summer School on Service-Oriented Computing, Hersonissos, Greece, 3–9 July 2022; Springer International Publishing: Cham, Switzerland; pp. 119–138.

24. Calancea, C.G. A review on Business-to-Business Data Sharing technical solutions. *Int. J. Optim. Appl.* **2022**, *2*, 37–41.

25. Sober, M.; Scaffino, G.; Schulte, S.; Kanhere, S.S. A blockchain-based IoT data marketplace. *Clust. Comput.* **2023**, *26*, 3523–3545. [CrossRef]

26. Meijers, J.; Putra, G.D.; Kotsialou, G.; Kanhere, S.S.; Veneris, A. Cost-effective blockchain-based iot data marketplaces with a credit invariant. In Proceedings of the 2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), Sydney, Australia, 3–6 May 2021; pp. 1–9.

27. Ullo, S.L.; Sinha, G.R. Advances in smart environment monitoring systems using IoT and sensors. *Sensors* **2020**, *20*, 3113. [CrossRef] [PubMed]

28. Bordel, B.; Alcarria, R.; de la Torre, G.; Carretero, I.; Robles, T. Increasing the Efficiency and Workers Wellbeing in the European Bakery Industry: An Industry 4.0 Case Study. In Proceedings of the International Conference on Information Technology & Systems, San Carlos, Costa Rica, 9–11 February 2022; Springer International Publishing: Cham, Switzerland, 2022; pp. 646–658.

29. Li, Y.N.; Feng, X.; Xie, J.; Feng, H.; Guan, Z.; Wu, Q. A decentralized and secure blockchain platform for open fair data trading. *Concurr. Comput. Pract. Exp.* **2020**, *32*, e5578. [CrossRef]

30. Travizano, M.; Sarraute, C.; Dolata, M.; French, A.M.; Treiblmaier, H. Wibson: A case study of a decentralized, privacy-preserving data marketplace. In *Blockchain and Distributed Ledger Technology Use Cases: Applications and Lessons Learned*; Springer International Publishing: Cham, Switzerland, 2020; pp. 149–170.

31. Nguyen, L.D.; Leyva-Mayorga, I.; Lewis, A.N.; Popovski, P. Modeling and analysis of data trading on blockchain-based market in IoT networks. *IEEE Internet Things J.* **2021**, *8*, 6487–6497. [CrossRef]

32. Bordel, B.; Lebigot, P.; Alcarria, R.; Robles, T. Digital food product traceability: Using blockchain in the international commerce. In *Digital Science*; Springer International Publishing: Cham, Switzerland, 2019; pp. 224–231.

33. Dai, W.; Dai, C.; Choo, K.K.R.; Cui, C.; Zou, D.; Jin, H. SDTE: A secure blockchain-based data trading ecosystem. *IEEE Trans. Inf. Forensics Secur.* **2019**, *15*, 725–737. [CrossRef]

34. Xiong, W.; Xiong, L. Smart contract based data trading mode using blockchain and machine learning. *IEEE Access* **2019**, *7*, 102331–102344. [CrossRef]

35. Sánchez, B.B.; Alcarria, R.; Valladares, T.R. A Data-Driven Blockchain-Based Marketplace to Promote Transparency and Accountability in City-Region Food Systems. In Proceedings of the 2023 18th Iberian Conference on Information Systems and Technologies (CISTI), Aveiro, Portugal, 20–23 June 2023; pp. 1–6.

36. Wu, Y.; Wu, Y.; Cimen, H.; Vasquez, J.C.; Guerrero, J.M. Towards collective energy Community: Potential roles of microgrid and blockchain to go beyond P2P energy trading. *Appl. Energy* **2022**, *314*, 119003. [CrossRef]

37. Bodkhe, U.; Tanwar, S.; Bhattacharya, P.; Kumar, N. Blockchain for precision irrigation: Opportunities and challenges. *Trans. Emerg. Telecommun. Technol.* **2022**, *33*, e4059. [CrossRef]

38. Xu, Y.; Ahokangas, P.; Yrjölä, S.; Koivumäki, T. The fifth archetype of electricity market: The blockchain marketplace. *Wirel. Netw.* **2021**, *27*, 4247–4263. [CrossRef]

39. Wang, Z.; Yang, L.; Wang, Q.; Liu, D.; Xu, Z.; Liu, S. ArtChain: Blockchain-enabled platform for art marketplace. In Proceedings of the 2019 IEEE International Conference on Blockchain (Blockchain), Atlanta, GA, USA, 14–17 July 2019; pp. 447–454.

40. Li, J.; Grintsvayg, A.; Kauffman, J.; Fleming, C. LBRY: A blockchain-based decentralized digital content marketplace. In Proceedings of the 2020 IEEE International Conference on Decentralized Applications and Infrastructures (DAPPS), Oxford, UK, 3–6 August 2020; pp. 42–51.

41. Leduc, G.; Kubler, S.; Georges, J.P. Innovative blockchain-based farming marketplace and smart contract performance evaluation. *J. Clean. Prod.* **2021**, *306*, 127055. [CrossRef]

42. Ha, M.; Kwon, S.; Lee, Y.J.; Shim, Y.; Kim, J. Where WTS meets WTB: A Blockchain-based Marketplace for Digital Me to trade users' private data. *Pervasive Mob. Comput.* **2019**, *59*, 101078. [CrossRef]

43. Chang, Y.W.; Lin, K.P.; Shen, C.Y. Blockchain technology for e-marketplace. In Proceedings of the 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), Kyoto, Japan, 11–15 March 2019; pp. 429–430.

44. Kabi, O.R.; Franqueira, V.N. Blockchain-based distributed marketplace. In Proceedings of the International Conference on Business Information Systems, Berlin, Germany, 18–20 July 2018; Springer: Cham, Switzerland, 2018; pp. 197–210.

45. Treleaven, P.; Barnett, J.; Knight, A.; Serrano, W. Real estate data marketplace. *AI Ethics* **2021**, *1*, 445–462. [CrossRef]

46. Ranganthan, V.P.; Dantu, R.; Paul, A.; Mears, P.; Morozov, K. A decentralized marketplace application on thethereumm blockchain. In Proceedings of the 2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC), Philadelphia, PA, USA, 18–20 October 2018; pp. 90–97.

47. Bajoudah, S.; Dong, C.; Missier, P. Toward a decentralized, trust-less marketplace for brokered IoT data trading using blockchain. In Proceedings of the 2019 IEEE International Conference on Blockchain (Blockchain), Atlanta, GA, USA, 14–17 July 2019; pp. 339–346.

48. Malik, J.A.; Saleem, M. Blockchain and Cyber-Physical System for Security Engineering in the Smart Industry. In *Security Engineering for Embedded and Cyber-Physical Systems*; CRC Press: Boca Raton, FL, USA, 2022; pp. 51–70.

49. Faisal, T.; Dohler, M.; Mangiante, S.; Lopez, D.R. BEAT: Blockchain-enabled accountable and transparent network sharing in 6G. *IEEE Commun. Mag.* **2022**, *60*, 52–56. [CrossRef]

50. Bordel, B.; Alcarria, R.; Robles, T. Denial of Chain: Evaluation and prediction of a novel cyberattack in Blockchain-supported systems. *Future Gener. Comput. Syst.* **2021**, *116*, 426–439. [CrossRef]

51. Bauer, D.P. ERC-20: Fungible Tokens. In *Getting Started with Ethereum: A Step-by-Step Guide to Becoming a Blockchain Developer*; Apress: Berkeley, CA, USA, 2022.

52. Bradić, S.; Delija, D.; Sirovatka, G.; Žagar, M. Creating own NFT token using erc721 standard and solidity programming language. In Proceedings of the 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 23–27 May 2022.

53. Reitwießner, C.; Johnson, N.; Vogelsteller, F.; Baylina, J.; Feldmeier, K.; Entriken, W. EIP-165: ERC-165 Standard Interface Detection: Ethereum Improvement Proposals. Available online: https://eips.ethereum.org/EIPS/eip-165 (accessed on 30 March 2024).

54. Agarwal, U.; Singh, K.; Verma, R. An overview of non-fungible tokens (NFT). *Int. J. Adv. Res. Sci. Commun. Technol. (IJARSCT)* **2022**, *1*, 237–240.

55. Fúster-Sabater, A.; Requena, V.; Cardell, S.D. An Efficient Algorithm to Compute the Linear Complexity of Binary Sequences. *Mathematics* **2022**, *10*, 794. [CrossRef]

56. López, A.B.O.; Encinas, L.H.; Muñoz, A.M.; Vitini, F.M. A lightweight pseudorandom number generator for securing the Internet of Things. *IEEE Access* **2017**, *5*, 27800–27806. [CrossRef]

57. Bordel, B.; Orúe, A.B.; Alcarria, R.; Sánchez-De-Rivera, D. An intra-slice security solution for emerging 5G networks based on pseudo-random number generators. *IEEE Access* **2018**, *6*, 16149–16164. [CrossRef]

58. Bordel, B.; Alcarria, R.; Robles, T.; de la Torre, G.; Carretero, I. Digital user-industry interactions and Industry 4.0 services to improve customers' experience and satisfaction in the European bakery sector. In Proceedings of the 2021 16th Iberian Conference on Information Systems and Technologies (CISTI), Chaves, Portugal, 23–26 June 2021; pp. 1–10.