


Managing Digital Objects with Decentralised Identifiers based on NFT-like schema

Chunming Rong , Jiahui Geng, Martin Gilje Jaatun*

*University of Stavanger, Norway {chunming.rong,jiahui.geng,martin.g.jaatun} @uis.no

Abstract—The diversity (text, images, algorithms, etc.) and the ambiguity of data sovereignty and privacy make the management of digital objects very challenging. Users need a unified and convenient way to manage their digital objects. This places a high demand on the findability and interoperability of the management model. In recent years blockchain has provided a new route to an open and secure platform due to its attributes such as distributed, traceable, and tamper-evident. In this paper, a new NFT-like scheme is proposed, which uses metadata converts digital assets into digital object identifiers, and transforms digital objects that require clear sovereignty into NFTs to ensure the authenticity and uniqueness of ownership. Our scheme can facilitate the dynamic management of digital objects using smart contracts.

Index Terms—Digital Object Management, Metadata, DOI, DID, NFT

I. INTRODUCTION

With the widespread use of artificial intelligence, mobile Internet, cloud computing, 5G and other new-generation information technologies in various industries, people are accelerating the process of datafication, producing massive digital objects all the time. Now the importance of data has been recognized, and business data has become the core production factor of enterprises. With the help of artificial intelligence technology, production efficiency can be improved. And personal data can also help companies identify potential users.

However, data objects are very easy to copy and spread, and in the early days of the Internet era, the protection of data property rights was ignored, making the management of digital objects in a very chaotic state. The misuse of personal privacy data, unclear definition of data property rights and data silos highlight the immaturity in the process of production, collection, analysis and transaction. Data object management and the data object market are still in the early stage and need to be nurtured by the simultaneous development of legal systems and technologies.

The United States and the European Union passed legislation in 2019 to promote the open sharing of government data, emphasizing open data in a standardized, machine-readable form and exploring real-time data sharing through API interfaces [1], [2]. In terms of data property rights and protection, the EU's General Data Protection Regulation (GDPR) [3] provides a detailed definition of personal data property rights. China also proposed the Data Security Law [4] and the Personal Information Protection Law [5], which laid a preliminary legal foundation for regulating the data element

market. The protection of data sovereignty is particularly critical in the context of geopolitics, which challenges the digital economy's globalization. Establishing an ownership or intellectual property protection system for digital objects requires consideration of multiple parties' interests along each data pipeline; it emphasizes the balance of data access, control, and rights and interests distribution.

Digital Object Identifier (DOI) system is currently an essential tool for digital object management. It includes a mechanism for transferring resources into unique, persistent identifiers and a protocol for resolving the identifiers into the addresses where the resources reside. DOI was designed to overcome the disadvantage that URLs change frequently and do not display information about the digital resources themselves. However, the current DOI system is centralized. Although DOIs are content-based, the definition of the content is entirely up to the DOI registration at the publisher level. Besides, just like registering a domain name, registering a digital object as a DOI still requires payment to the publisher. The publisher needs to pay a substantial annual fee to CrossRef [6], an official DOI registration agency. Such a centralized mechanism will inevitably lead to platform monopoly, and people will worry that the centralized platform will take away potential profits. Once the cost of registering digital assets is higher than the benefits provided by the DOI resolution system itself. This in turn inhibits sharing and circulation of data identities.

In recent years, researchers have proposed a new open-source initiative, OpenIaC, the network is my computer [7], which organizes the digital objects distributed in the network into a form that can be easily accessed, managed, and shared through efficient management tools. Digital objects are not unique. They are naturally susceptible to tampering or copying, and they are vulnerable to attack and destruction during production and trading, thus violating the value of the digital objects themselves. NFT is a blockchain-powered cryptographic digital certificate used to record the ownership of virtual digital assets, such as artworks or collectibles. It is unique, irreplaceable, indivisible, programmable, traceable, and permanently preserved. Inspired by this, we propose using decentralized identities to create identities for important distributed digital objects in the network, combined with the mechanism of NFT for sovereign registration. At the same time, our scheme utilizes metadata to provide enhanced discoverability and interoperability of data. By writing smart

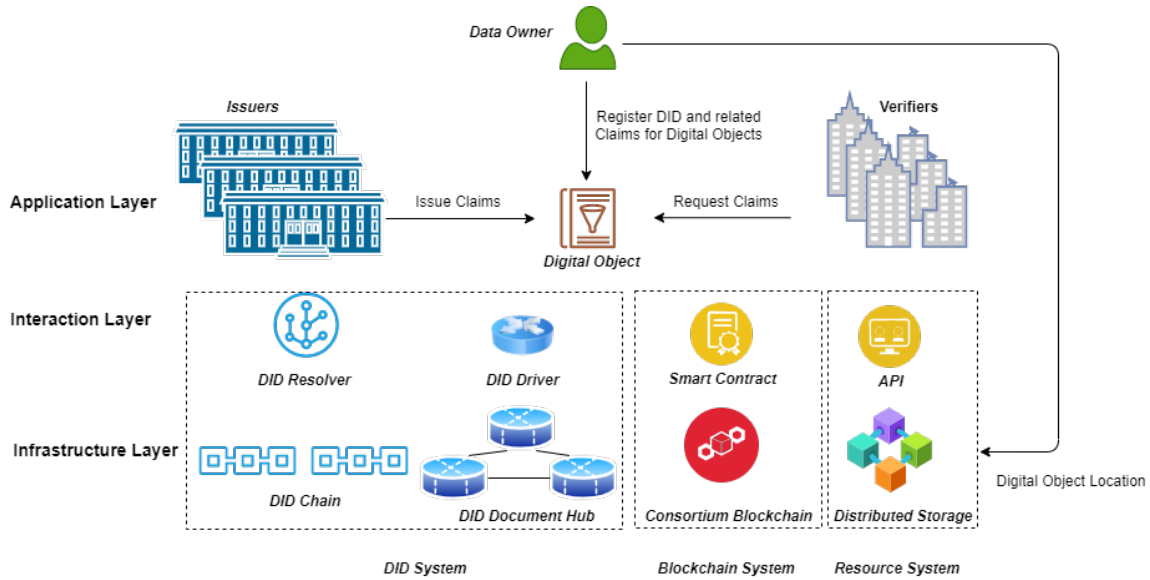


Fig. 1. Architecture Design for digital object management.

contracts to access the data, the system can automatically and precisely complete some digital object management operations.

A. Contribution

The contributions of this paper are summarized as follows:

- We propose to replace the traditional DOI with DID as the identifier management system for decentralized distributed digital objects.
- We propose to use NFT-like schema to represent ownership of digital objects, and use smart contracts to handle other rights of digital objects flexibly.
- Metadata-based DID for enhanced retrievability and interoperability of digital objects Our solution can be integrated with smart contracts to automate data management and access.
- Our work will inspire the implementation of Open Infrastructure and help enable the use but not disclosure of data.

B. Organization

The remainder of this paper is structured as follows. Section II describes the challenges and the principles for digital object management. In Section III, we introduce related work. Section IV discusses the design ideas and our proposed approaches. Section VI concludes the paper.

II. CHALLENGES AND PRINCIPLE OF DIGITAL OBJECT MANAGEMENT

Digital object management faces many challenges from the data characteristic and the data flow process.

- Complex ownership of data
Take the data generated by cell phone applications as an example. For application developers, user data is fundamental for improving product performance and mining

business opportunities, and their services are based on user data. Cell phone manufacturers also need to make full use of user data to improve the ease of use, versatility, and cross-device collaboration of cell phones. In fact, users, as data producers, have a minor voice because they lack the channel to express their opinions and are not able to monitor and determine how their data is used.

- The value of data is hard to measure
The value of data includes, but is not limited to:

- 1) The cost of acquiring and storing data.
- 2) The loss that would result from data loss
- 3) The cost of mitigating potential risks associated with data.
- 4) The potential sales value of the data in the marketplace.

There is no uniform standard for quantifying data value, and the value of data is contextual and depends on how the data is used and the use case. A small amount of data is of little value, and the data collected at scale contains an enormous value.

- Data diversity, scale and time effect
The data sources are very diverse, and may come from personal, commercial, or public government activity. These data are collected, stored, processed, and analyzed to form derived data and related information knowledge and algorithms. Data has some degree of substitutability. For example, in machine learning algorithms, the lack of some sample data or features does not have a decisive impact on the model performance. The data itself is time-sensitive, and the value of existing data will slowly decrease over time.

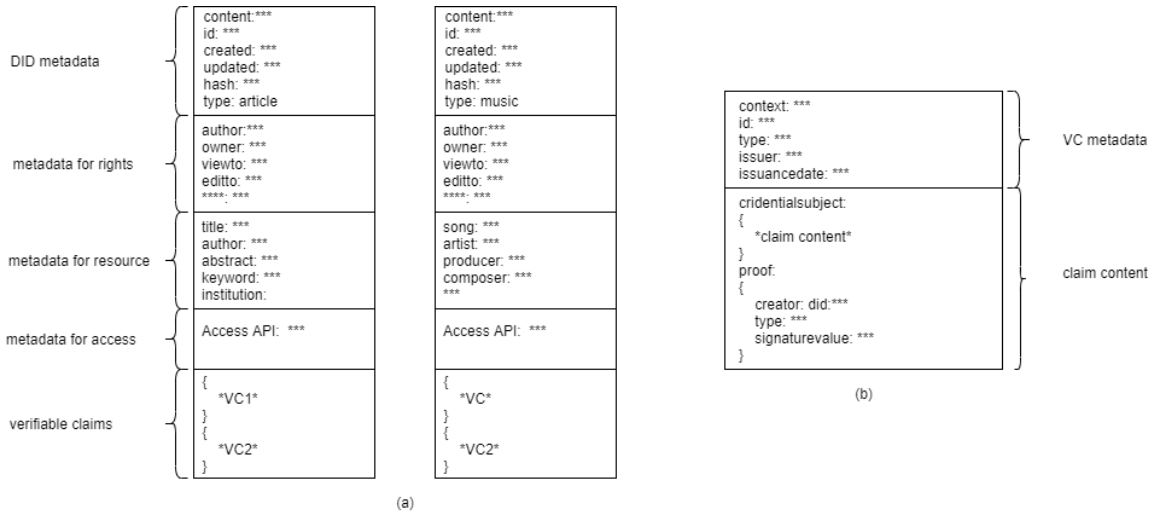


Fig. 2. Design of the DID Document for digital objects

III. RELATED WORK

A. Digital Object Identifier

Digital Object Identifier(DOI) is an internationally accepted identifier for digital resources. It is widely used to identify academic, professional, and government resources, such as journal articles, research reports, datasets, and other publications, as well as other types of information resources, such as commercial videos. DOI system provides DOI registration, resolution, management, and a variety of value-added services [8]. The DOI system was officially released as an ISO international standard in 2012 and has become a global standard for digital resource identification and linking. Unlike ISBN or ISRC [9], which are only used as identifiers, DOI system uses the indecs content model [10] to represent metadata, and achieves digital object management with interoperability by binding related metadata.

B. Blockchain and Smart Contract

Blockchain technology was introduced in the decentralized cryptocurrency system Bitcoin [11]. Blockchain is essentially an advanced distributed database that allows information to be shared transparently across enterprise networks. Blockchain technology uses block and chain data structure to verify and store data, distributed node formula algorithms to generate and update data, and cryptography to ensure the security of data transmission and access.

The smart contract is an automatically executed contract that contains code and protocols running in a distributed, decentralized blockchain network. The code controls execution, and transactions are traceable and irreversible. Smart contracts allow trusted transactions and agreements between different anonymous parties without needing a central authority, legal system, or external enforcement mechanism. Smart contracts are data transparent, tamper-evident and perpetual.

C. Decentralized Identifier and Verifiable Credentials

A Decentralized Identifier (DID) is a new type of identifier that is globally unique, resolvable, and cryptographically verifiable. The W3C has proposed the concept of a blockchain-based distributed digital identity solution that allows users to generate and control their own digital identity without relying on a specific service provider. When an identity is created, corresponding encryption keys (public and private keys) are generated. The identity wallet submits a registration payload with the public key to the blockchain, and the blockchain generates a unique identifier for the identity wallet. The private key remains on the user's device or identity wallet and is used during authentication. DID can be based on biometrics [12] or traditional identification documents [13].

Verifiable credentials(VC) are an open standard for digital credentials that can represent information found in physical credentials, such as passports or licenses [14]. They have many advantages over physical credentials, for example they are digitally signed, which makes them tamper-proof and instant verification. The public infrastructure required for verifiable credentials is provided by DID, where each issuer, subject and verifier creates a unique identifier and associates a set of public keys with their identifiers. The issuer's public key is public, so any verifier can verify the verifiable credentials generated by the issuer.

D. Non-Fungible Token

Fungibility is the ability that a good can be readily interchanged with another one of same or similar kind. Like the physical currencies, cryptocurrencies are fungible, meaning they can be traded or exchanged. This fungibility feature makes cryptocurrencies suitable as a secure medium of exchange in the digital economy. NFTs change the paradigm of cryptocurrencies by making each token unique and irreplaceable, thus making it impossible for one irreplaceable token to be equivalent to another. NFTs are digital assets

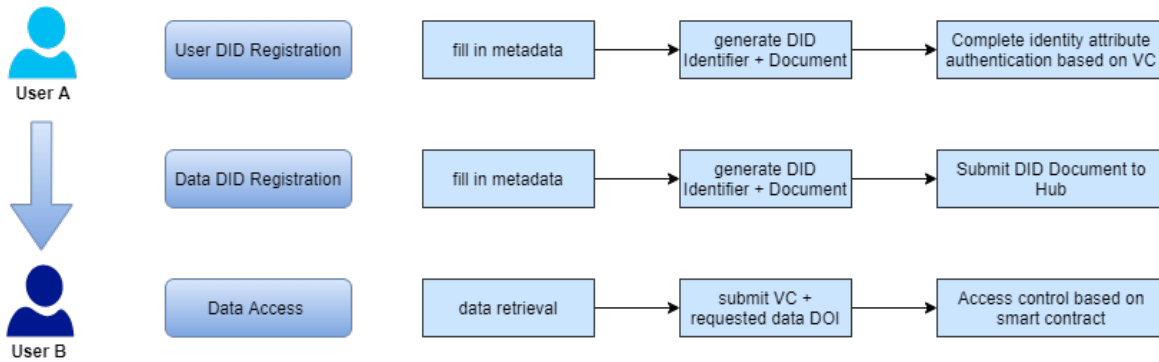


Fig. 3. The proposed workflow of our platform

created, maintained and executed by smart contracts. They are deployed on blockchains such as Ethereum [15] through standard contract forms such as smart contracts ERC-721 [16] and ERC-1155 [17]. NFTs rose to fame in 2017 with a game called "CryptoKitties" [18], and since then, game developers have heavily adopted NFTs to allow gamers to win in-game items such as digital gear or prizes, as well as other game collectibles. In addition to games, NFTs are often used to sell a variety of virtual collectibles, including NBA virtual trading cards, music, digital images, video clips, and even virtual real estate in the virtual world [19].

IV. PROPOSED APPROACH

A. Architecture Design

The system is designed to register a unified digital identifier for digital objects stored in the resource system to enhance the findability and interoperability of digital resources. It will also ensure secure and controlled access to data resources. Our decentralized solution is illustrated in Figure 3. In general, the system consists of three layers. The top layer is the application layer, which includes different organizations like Issuers, Verifiers, and the subject of the article: digital objects. Digital objects generally refer to images, audio, video, PDF, etc., but can also be large data sets or models. The middle layer is the interaction layer, which is responsible for the interaction between application layer and infrastructure layer, including DID resolution, data access-related APIs, and metadata-based smart contracts that handle simple logical operations, such as data access control, data integrity checking, and digital signature verification. The bottom layer is the infrastructure layer, which includes the DID system, blockchain system, and resource system. The resource system can also be referred to as the file system or storage system.

1) *Resource System*: In common scenarios, file sharing is generally local, peer-to-peer, rather than broadcast to everyone. It would be overwhelming to let the blockchain store massive amounts of data indiscriminately. Thus, it is reasonable to instead calculate the digital fingerprint of the file (MD5 or HASH) and upload it on the chain along with some other optional information. The data objects themselves are stored in private data servers, cloud storage spaces or InterPlanetary

File System(IPFS) systems [20]. If the security level of the file system is high, then the storage method such as IPFS should be avoided, and the private storage method should be preferred.

2) *Blockchain System*: The blockchain system mainly affects the entire system through smart contracts. It will complete simple logical operations based on the metadata in the DID system, including data integrity inspection during digital object registration and data access during data sharing and use. control, digital signature verification in DID documents, etc.

3) *DID System*:

- **DID Identifier**

The DID Identifier is a three-part string that contains the DID schema, DID method, and DID method-specific string. DID Method is used to distinguish how each DID is parsed. Different DID Chains have their own DID methods.

- **DID Chain**

The DID Chain serves as a public infrastructure on which the public keys of all entities within the network are stored. Based on distributed ledger technology, DID Chain will be used to verify the integrity of DID documents and enhance the trust and interoperability of the network.

- **DID Resolution**

DID Resolution mainly contains DID Resolver and DID Driver. The role of DID Resolver is to get the DID document by parsing DID Identifier so that the metadata of the data object can be obtained. The specification of DID Resolver is mainly based on DIF. DID Driver is DID Chain specific and is mainly responsible for converting machine readable information into human readable information. DIDs may exist on different blockchains, so a DID Resolver can be considered as the aggregator of DID Drivers.

- **DID Document and DID Document Hub**

A DID document is a document pointed to by the DID Identifier after it has been parsed. It contains the metadata describing the digital objects. In our protocol, the designed DID document is shown in Figure 2. A DID document mainly includes 1) DID metadata for describing the DID document itself, such as creation

time, update time, version number, data type, and digital signature for the whole DID document etc. 2) metadata for rights, used to describe the rights of digital objects, including author, owner, view right, edit right, etc. 3) metadata for resource, mainly describes the resource type, when the data type is article, metadata for resource will include title, author, abstract, keyword, institution and other basic information. When the data type is music, the metadata will include song title, artist name, producer name, composer name, writer name and so on. 4) metadata for access, 5) verifiable claims will mainly include claims obtained from other organizations or between data objects. For example, the description of the music product from the copyright is certified by an authority, or if the data object is a dataset, the claims can also be the measured performance of another algorithmic model data object on the dataset. The DID Document Hub is where the DID Document is stored.

B. Function Design

1) *DID Registration and Usage*: When a user wants to use our system to register DID Identifier for his data resources with copyright protection and limited sharing requirements, the user first needs to use a user agent, such as a mobile phone or a web page, and the user needs to fill in the metadata registration form. The metadata registration form is like We designed the same in DID Document. The user needs to accurately register the data information including the API interface accessed. After the registration of the form is completed, the system uses the smart contract to check the accessibility of the interface and the integrity of the data. Once the verification is successful, the system will generate a unique DID Identifier and Document pair based on the metadata data. The DID Identifier will be returned to the user, and the DID Document will be saved in the DID Document Hub.

DID Document Hub is a publicly accessible system, unlike other systems by design. In a system that authenticates users, DID Documents cannot be retrieved, and the entire DID system may be designed for selective disclosure of users only when the user actively presents the corresponding verifiable claims. Verifier can check the data in DID Documents then. However, in our system, since DID Document is based on metadata, and the metadata is public and retrievable.

2) *Data Access Control*: Since we have defined data access rights and data access API in the metadata, the system will implement smart contract based data access control. The data rights will be subdivided into access rights and editing rights, etc. The smart contract will first check whether the user's identity meets the requirements when it receives the data access request from the user, and if it passes, the smart contract will return the access API.

3) *Data Integrity Control*: The basis of data integrity is the DID Chain, which contains the public keys of all registered users and registered resources. Based on the blockchain's tamper-evident, open and transparent, and traceability features, DID Chain is a public ledger for verifying the validity of all

signatures in DID Documents. Based on the public key of the creator, we can verify the integrity of DID Document by verifying the signatures in DID metadata. Based on the digital signatures in the proof in Verifiable Claims, we can verify the authenticity of Verifiable Claims.

4) *FAIR Principle Compliance*: Our system assigns globally unique and persistent identifiers to metadata. We use rich metadata to describe the data, and in the DID Document Hub, the metadata in the documents are registered and indexed, and can be retrieved in searchable engines. In addition, the Variable Claims section of our data is designed to include references to other metadata and metadata, reflecting the interoperability of our system. In summary, our system fully satisfies the FAIR data management principals.

C. Workflow

Figure 3 shows our workflow. First, a decentralized identity needs to be registered for users in the system. The user first needs to fill in the metadata to generate the DID Identifier and DID Document. The user's DID Document only records the user's basic information, and the user needs to authenticate more attributes through the VC system. These properties will help users gain access to specific digital objects. For example, he can get a VC from his employer to access the resources.

Then the user needs to create a DID for the digital object, and the user needs to fill in the metadata of the data, as shown in Figure 2, which includes information for findability and interoperability, access methods, and access policies. After the corresponding DID Identifier and Document pair are generated, the DID Document will be transferred to the DID Document Hub, where a retrieval service and authentication and access control smart contract based on metadata information is provided.

When a user needs to access data, he can directly obtain the corresponding digital object DOI from other users or retrieve the required data from the provided service. The user needs to provide the corresponding VC and DOI of the corresponding digital object. The verification process is automatically performed by the smart contract. When the verification is passed, the user will directly access the data, otherwise the service will stop.

V. DISCUSSION

Table I shows the comparison of our method with traditional DOI methods. 1) The DOI approach is centralized and it relies on Crossref, an institution led by the International DOI Foundation. Our solution is decentralized, or federated. This is determined by how the underlying service operates. 2) The parsing of DOI is static, while our smart contract-based solution is dynamic, making it easy to adjust and upgrade. 3) The DOI scheme is only responsible for maintaining the metadata and references hosted on its site, while our scheme combines metadata, API and smart contracts. 4) The DOI scheme does not provide follow FAIR principle, while our scheme satisfies. 5) The trust source of the DOI is some authoritative publishers and web service providers, and our

TABLE I
COMPARISON BETWEEN TRADITIONAL DOI AND OUR SOLUTION

	DOI	DOID
Architecture	centralized	decentralized/federated
Interpretation	static	synamic
Components	metadata+reference	metadata+API+smart contract
Access Control	No	fine-grained access control use but not disclose
FAIR principle	No	Yes
Trust Source	publisher/web service provider	blockchain
Support for Market Trading	No	Yes
Support for Open Infrastructure	No	Yes

scheme is based on the immutability of the blockchain. 6) DOI does not support data object market transactions, and our method uses NFT schema to support market trading. 7) DOI does not support Open Infrastructure, but our method supports it [7].

VI. SUMMARY

Blockchain technology has been widely studied and applied in recent years to build decentralized cooperative networks. Based on peer-to-peer networks, consensus mechanisms and cryptography, blockchain naturally has the advantages of openness, security and trustworthiness. Technologies such as smart contracts, decentralized identifiers, and unforgeable tokens have greatly expanded the application scenarios of blockchain. In this paper, we propose a decentralized approach to solve the problem of closure and monopoly in data object management. We propose to use metadata-based DIDs to construct identities for data objects and Verifiable Credentials as an authentication method for interactions between data. Our solution enhances data management capabilities, improves data value, and protects users' data rights. Our approach lays the foundation for implementing NFT-like open data asset transactions.

ACKNOWLEDGMENT

REFERENCES

- [1] "Open government data act," <https://www.cio.gov/handbook/it-laws/ogda/>, accessed: 2022-08-02.
- [2] "Directive (eu) 2019/1024 of the european parliament and of the council of 20 june 2019 on open data and the re-use of public sector information," <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024>, accessed: 2022-08-02.
- [3] P. Voigt and A. Von dem Bussche, "The eu general data protection regulation (gdpr)," *A Practical Guide, 1st Ed.*, Cham: Springer International Publishing, vol. 10, no. 3152676, pp. 10–5555, 2017.
- [4] "Data security law of the people's republic of china," <http://www.npc.gov.cn/englishnpc/c23934/202112/1abd8829788946ecab270e469b13c39c.shtml>, accessed: 2022-08-02.
- [5] "Personal information protection law of the people's republic of china," https://www.pcpd.org.hk/english/data_privacy_law/mainland_law/mainland_law.html, accessed: 2022-08-02.
- [6] "Cryptokitties," <https://www.crossref.org/>, accessed: 2022-08-02.
- [7] C. Rong, J. Geng, T. J. Hacker, H. Bryhni, and M. G. Jaatun, "Openiac: open infrastructure as code-the network is my computer," *Journal of Cloud Computing*, vol. 11, no. 1, pp. 1–13, 2022.
- [8] N. Paskin and I. D. Foundation, *The DOI@ handbook*. Citeseer, 2002.
- [9] "Isrc handbook," https://www.ifpi.org/wp-content/uploads/2021/02/ISRC_Handbook.pdf#Heading321, accessed: 2022-08-02.
- [10] "The indecs_i metadata framework: Principles, model and data dictionary," https://www.doi.org/topics/indecs/indecs_framework_2000.pdf, accessed: 2022-08-02.

- [11] S. Nakamoto and A. Bitcoin, "A peer-to-peer electronic cash system," *Bitcoin.–URL: https://bitcoin.org/bitcoin.pdf*, vol. 4, p. 2, 2008.
- [12] A. Othman and J. Callahan, "The horcrux protocol: a method for decentralized biometric-based self-sovereign identity," in *2018 international joint conference on neural networks (IJCNN)*. IEEE, 2018, pp. 1–7.
- [13] J. Geng, N. Kanwal, M. G. Jaatun, and C. Rong, "Did-efed: Facilitating federated learning as a service with decentralized identities," in *Evaluation and Assessment in Software Engineering, 2021*, pp. 329–335.
- [14] "Verifiable credentials," <https://verifiablecredential.io/learn>, accessed: 2022-08-02.
- [15] G. Wood *et al.*, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum project yellow paper*, vol. 151, no. 2014, pp. 1–32, 2014.
- [16] "Erc-721," <https://ethereum.org/en/developers/docs/standards/tokens/erc-721/>, accessed: 2022-08-02.
- [17] "Erc-1155," <https://eips.ethereum.org/EIPS/eip-1155/>, accessed: 2022-08-02.
- [18] "Cryptokitties," <https://www.cryptokitties.co/>, accessed: 2022-08-02.
- [19] "Welcome to decentraland," <https://decentraland.org/>, accessed: 2022-08-02.
- [20] "Pfs powers the distributed web," <https://ipfs.tech/>, accessed: 2022-08-02.