A blockchain-based unified PID with applications in healthcare

Jose Armando Hernandez Gonzalez, Miguel Colom

A blockchain-based unified PID with applications in healthcare

Jose Armando Hernandez
Centre Borelli, ENS Paris-Saclay
Gif-sur-Yvette, France
jose.hernandez_gonzalez@ens-paris-saclay.fr
0000-0002-6692-8640

Miguel Colom
Centre Borelli, ENS Paris-Saclay
Gif-sur-Yvette, France
miguel.colom-barco@ens-paris-saclay.fr
0000-0003-2636-0656

***Abstract*—This work proposes DOIschain, a decentralized blockchain-based architecture to generate and store persistent, immutable, and perennial identifiers for digital objects (IDO) for the final products of computational scientific research. A supra identifier is proposed to guarantee the global reproducibility, traceability, authenticity, and provenance of the results and artifacts of scientific research. Reproducibility and credibility are achieved through the correct, reliable, and FAIR-compliant identification in blockchain and preservation in the Interplanetary File system (IPFS) of the computational artifacts, such as metadata, the publication paper, the software source code, and datasets. A use case in the health sector is presented within the framework of the BraTS (Brain Tumor Segmentation) competition and a simplified proof-of-concept DOIschains implementation where a persistent generic basic PID for medical datasets is described.**

***Index Terms*—Persistent Identifier, Blockchain, IPFS, medical datasets**

I. Introduction

The preservation, hierarchization, ordering, classification, and recovery of the location of the information needs of persistent identifiers (PIDs). These are conceived as a long-term reference to objects.

PIDs are used to archive, reference, describe, and cite [1], and their use has been widely adopted by journals and data repositories worldwide [2]. PIDs allow different platforms to interact and exchange information through standard PID formats, thus allowing for FAIR services, simplified workflows, and databases that can be queried with standard identifiers as, for example, those proposed by DataCite [3] and Crossref[1]. However, each repository or database generally implements its own non-standard identifiers, typically because of the myriad of sources and the many of information management software available.

Some PIDs are specialized and tailored depending on the application or research field, and some repositories and databases developed their own PID formats because of their own information management purposes. We can find intrinsic (computed using the binary contents of the object itself) and extrinsic (assigned by an authority) PIDs with different granularity such as the swMATH-ID (mathematics), the SHWID (software), the DOI (journals), the ORCID (researchers) [2], or the MeSH (National Library of Medicine's Medical Subject Headings), to cite a few examples.

In computational scientific research, three fundamental elements are identified: the article, the software, and any associated datasets. Indeed, to verify the claims of the article, one needs to execute the same version of the software used by the article's authors with exactly the same input data. In this sense, much effort has been put into developing PIDs for the article (for example, with the DOI or with repositories such as ArXiv and others). Software and datasets have been treated interchangeably in the FORCE Data Citation initiative focused on software with identifiers such as SWHID or SWMATH-ID [4], for example. However, the adequate persistent identification and versioning of the data have not been properly or sufficiently well addressed. According to its particular characteristics, unlike software, the data requires the availability of large storage resources. Existing data repositories do not have standard PIDs adapted to the specifics and needs of research datasets.

Many challenges, for example, BraTS, Kaggle, or ILSVRC, and benchmarks, strongly rely on well-known datasets. We can cite for example COCO[2], ImageNet[3] and the BraTS dataset[4] public dataset repositories. However, in many cases, these datasets are maintained and versioned privately without following any known standards. When they are cited in publications through their URL, it might happen that the object disappears when the maintaining entity decides. Typically, the decision is related to the cost of the infrastructure that makes them available for the long term.

The use of Content Identifiers (CIDs) and the InterPlanetary File System (IPFS) can be a solution for the problem of centralized datasets and their availability [5]. Inaccessible data because of broken URLs or incorrectly inventoried compromises reproducibility, FAIR compliance, and general trustworthiness

The authors would like to thank the financial support to the SESAME's *OVD-SaaS* project from Région Île de France and BPI France. Also, to the Ministry of Science, Technology and Innovation of Colombia (Minciencias), within the framework of Call 885 of 2020, to finance the Doctorates of Excellence Program.

[1]https://www.crossref.org/

[2]https://cocodataset.org/#home

[3]https://www.image-net.org/

[4]https://www.kaggle.com/datasets/awsaf49/brats2020-training-data

of results.

This work focuses on PIDs for static digital objects, artifacts that, once created, do not change and need to be available long-term to ensure the reproducibility and integrity of research works that utilize them.

Each of the three elements of a reproducible publication has its specific role. The article presents the research claims and explains the method or algorithm in detail. The software becomes the actual verifiable implementation of the method. The datasets and metadata constitute the input and setup of the algorithms.

The data presents its own difficulties as it generally does not benefit from standard PIDs and might be stored in centralized repositories. Although there are already recommendations for the management and publication of datasets[5], the combined use of intrinsic PIDs and CIDs IPFS in a blockchain with a *supra-identifier* that encompasses all the three mentioned components has not been yet considered.

Our proposal to address the problem of centralized non-permanent objects in datasets associated to scientific publications is a follows.

1) A generic basic PID specialized in data on IPFS storage, specifically for medicine, is briefly described.
2) Create a DOIschain supra-identifier based on codemeta.json[6], including also the identifier associated to the article (the DOI), the software (SWHID), and the dataset PID.
3) Implement a private blockchain with proof of authority (PoA) via a consensus algorithm to record the transaction of generation of DOIschains. This allows the validation of their veracity and authenticity before the scientific community, as authorized nodes must create these DOIschains. The transaction is saved and validated on the blockchain ledger, and IPFS saves the metadata of the DOIschains. Both the metadata and datasets with intrinsic PIDs CID are stored in IPFS.

The article is organized as follows. Section II reviews the state of the art and compares the existing PID strategies with our proposal on blockchain-based PIDs. Section III describes a basic PID for medical datasets and how Blockchain is used to issue a supra PID DOIschains. Section IV discusses the reach and limitations of our proposal around an actual prototype that we built, which can be considered as a reference implementation. Section V presents how our system behaved in a real case, such as the BraTS competition. Finally, Section VI concludes the paper and presents ideas for future work.

II. State of the art on PIDs

Existing PIDs for People, Organizations, Data, and Publications can be classified according to their type and purpose [6]. According to the type of object they point to, we can classify them into three large groups: For People and Organizations, For Publications, and General Uniform Resource Identifiers (URIs).

Examples of type (1) are the Open Researcher and Contributor ID (ORCID)[7] and the Research Organization Registry (ROR)[8]. For type (2), the Virtual International Authority File (VIAF)[9], the International Standard Name Identifier ISNI)[10], or the International Standard Book Number (ISBN)[11]. Finally, URIs point to resources on the Web, each web page or file having a unique Uniform Resource Locator (URL). Finally, for type (3) the most representative would be the Archival Resource Key (ARK)[12] with 8.2 billion ARKs issued, the Handle System Digital Object Identifier (DOI)[13] with 200 million DOIs issued, the Magnet link (decentralized, with BitTorrent), the Uniform Resource Names (URNs)[14], the Extensible Resource Identifiers (XRIs)[15], the Persistent Uniform Resource Locators (PURLs), the Software Heritage identifiers SWHIDs)[16] [7], the Wikidata Identifier QIDs[17] or the UUIDs (Universally Unique IDentifier) -also known as GUIDs (Globally Unique IDentifier)- [8], just to cite some of the most relevant ones.

The DOI Foundation is a not-for-profit organization that governs the Digital Object Identifier (DOI) system on behalf of the agencies that manage DOI registries and their registration authority for the ISO 26324 standard.

The advances in computing technology in the last decades and the increase in the number of scientific articles being published every year (that need to cite data) become a reasonable explanation for the proliferation of data repositories, especially within the Confederation of Open Access Repositories (COAR) [18] that issued Community Framework for Good Practices in Repositories [9] since they utilize PIDs for discoverability purposes.

An essential attribute of DOIs is whether they are extrinsic or intrinsic. Extrinsic DOIs are generated and attributed by some external authority. In contrast, intrinsic identifiers are computed considering the contents of the digital object itself (typically, by considering its cryptographic hash).

For example, the Zenodo[19] repository is a well-known repository to store and check data. It assigns DOIs to reference digital objects of different nature, including software, data, and other type of documents. Zenodo stores large volumes of information. As an intrinsic identifier specific to source code, we can cite the Software Heritages project to preserve all of

[5]https://www.nature.com/sdata/policies/repositories
[6]https://codemeta.github.io/codemeta-generator/
[7]https://orcid.org/
[8]https://ror.org/
[9]https://viaf.org/
[10]https://isni.org/
[11]https://www.isbn-international.org/
[12]https://arks.org/
[13]https://www.doi.org/
[14]https://www.iana.org/
[15]https://www.oasis-open.org/
[16]https://www.softwareheritage.org/
[17]https://www.wikidata.org/wiki/Wikidata:Identifiers
[18]https://coar-repositories.org/
[19]https://zenodo.org/

the available software source code by permanently archiving it and indexing it with their SWHID PID[20].

Extrinsic identifiers are managed by authorities who take care of the issuance, ensure that the referenced objects exist and do not change, and revoke them when needed. These extrinsic PIDs are strongly centralized around an authority, and they can become invalid or even vanish if the authority removes them from its index for any reason. W3C recommends the use of Decentralized Identifiers (DIDs) [10], which come with wished benefits such as uniqueness, non-ambiguity, persistence, integrity, security, or gratuity [7].

Schaufelbuhl et al. propose in their prototype EUREKA [11] to use blockchain for the management of scientific publications, [12] document verification Doc-chain tool and scientific data reproducibility badging [13]. Although all those do not consider the issue of PIDs, these are relevant cases of blockchain used to solve problems related to referencing scientific publications. Similar to badging, blockchain-based CertAPP [14] has been proposed, and a new form of artifact traceability at the level of student projects [15] suggests the possibility of extending the solution to reproducible artifacts derived from computational scientific research. Recently, basic models of PIDs in blockchain have been proposed [16] such as T-PID, and other works based on IPFS [17].

Regarding the storage of medical datasets [18] [19], these analyze the benefits of using IPFS to store this type of medical data over traditional on-premise or cloud storage. Others propose IPFS and blockchain to specifically store and manage patient medical information [20], including images [21] [22], which depending on the number of patients treated can be quite large.

In the following, we shall present, as part of the review of the SOTA, the particular case of PIDs in biomedical sciences in section II and the blockchain-based PIDs in section III.

The particular case of PIDs in biomedical sciences

Currently, in several publications, a reasonable solution to the problem of identifiers is to assign Digital Object Identifiers (DOIs) to datasets. DOIs are already widely used by publishers as persistent identifiers for scholarly publications. They have been adopted by generalist data repositories such as Dryad, FigShare, Zenodo, Whole tale, and Dataverse, as well as by domain-specific repositories outside of biomedicine. The underlying DOI system, Handles, may also be used directly. The DataCite consortium provides a robust central mechanism for assigning DOIs to data.

However, DOIs are not commonly used for biomedical data, which is distributed across hundreds of autonomous, independently funded repositories. Instead, biomedicine has a longstanding practice of using a prefix combined with a locally assigned accession number as a unique identifier. Creating DOIs for the billions of existing entities identified via the prefix model may not be practical. Even if it became acceptable in the community, compatible with the traditional format of citing publications, financially feasible, and technically achievable, the tangible benefits could be significantly reduced by the complexity and cost of mapping original identifiers to the new DOIs. and it is forcing a PID system to handle and cite a type of information for which it has not been designed.

Mesh [23][21] is an example of a traditional PID that evolves and adapts to the advances of a domain as specific as medical science, considering that at that time it adapted to old physical library systems. There were no computer systems for large databases to archive and index information with comparable performance as nowadays. However, the latest advances in computer technology in relation to medical sciences have caused a strong dependence on data. So, this identifier does not support descriptions of medical software or datasets.

With the COVID-19 pandemic, several difficulties and limitations of the current identification of medical datasets became relevant. For the first time in this era of computer technology, the efforts of the entire scientific community were united in the fight against the virus. Many private medical databases, such as NCBI with their own indexing system, gave free access to information on virus genetic sequences, immunological epitopes[22], or tools of protein software (BLAST[23]). Other contributions surveyed and classified existing literature [24] and datasets available [25].

Much research on the topic was published, but however, the veracity and legitimacy of the results of many articles remain questionable due to defects in their reproducibility. Among them is the origin of the datasets that were used.

Some works have proposed metadata schemes for the discovery of medical datasets [26] from which the search platform DataMed [27] and OmicsDI[24] are searchable research and data indexes that emerged, bringing together several repositories along with the corresponding PIDs for each database. The Data Tag Suite (DATS) model supports the DataMed data discovery index. DataMed aims to do for data what PubMed has done for scientific literature. Similar to how the Journal Article Tag Suite (JATS) is utilized in PubMed, the DATS model facilitates the submission of dataset metadata to DataMed.

Although other works on the biomedical topic have worked on compact identifiers for biomedical data [28] and their citation via Uniform Resource Identifier (URIs), the use of IPFS is a modern approach that also imposes new challenges considering its distributed nature of this storage. Therefore, it is a pressing need to define a PID for medical data from scientific research in blockchain and IPFS according to good practices [29] and recommendations [6].

For illustrative purposes of the entire system, we propose a basic structure for a PID for medical datasets stored in IPFS, given that the existing PIDs are not well adapted to their use in the medical context and blockchain technologies, given that:

- Although DataMed is widely used, it is not a PID by itself, but rather a tool for finding datasets, it does

[20]https://www.swhid.org/

[21]https://www.nnlm.gov/guides/data-glossary/persistent-unique-identifier

[22]https://www.iedb.org/

[23]https://blast.ncbi.nlm.nih.gov/Blast.cgi

[24]https://www.omicsdi.org/

not mint identifiers for datasets, rather, it relies on the identifiers provided by the source [27]; no unified PID exists for medical datasets in blockchain technologies like MeSH is for medical literature.
- There is a need for PIDs for medical datasets that provide special crypto-token properties as uniqueness, immutability and non-fungibility (unique and indivisible digital identifier representing a numeric active that is recorded on a blockchain and is used to certify ownership and authenticity) in the IPFS cloud because it is not possible to publish the same dataset with the same hash-based CID.

The Joint Declaration of Data Citation Principles (JDDCP) [30] outlines core principles regarding the purpose, function, and attributes of data citations. The first of these principles is that data should be recognized as legitimate, citable research products, these principles rely heavily on the type and attributes of PID used to cite. Specifically, Principle 3 mandates that any research findings based on data must cite that data. Principle 4 requires that cited, archived data be assigned a globally unique, machine-resolvable persistent identifier, which should appear in the reference list of the citing article. The above implies that it must be resolvable. New technologies such as the Ethereum Name Service (ENS) [31] could be a good complement to the traditional DNS by changing the concept of a globally unique and resolvable machine that identifiers must resolve on the Web as HTTP Uniform Resource Identifiers (URIs), to comply with PID must be robust over time to changes in the underlying location of the data.

As an example, we briefly describe a PID MeSH-like, specifically tailored to reference biomedical data, which is a basic description for medical datasets PID based on blockchain and IPFS technologies. It is persistent, robust, reliable, and perennial for the very long term, and moreover, it follows the guidelines of the JDDCP.

III. Our proposal for a blockchain-based identifier: the *DOIschain*

As seen in the previous section II, there are open issues regarding PIDs. Blockchain systems for badging and managing scientific publications have been proposed [13], but a blockchain PID system for scientific publications based on computation extendable to different IPFS repositories has not been proposed. The goal is to improve aspects such as discovery, access, reuse, integrity, preservation, and FAIR access.

From a blockchain point of view, although issuing badges has certain similarities to issuing PIDs, the latter has more complexities associated with information management and Big Data 4V (volume, speed, variety, and veracity). A data science research project has three main elements (publications, algorithms/methods/code, and the datasets), which represent a large amount of information that must be correctly identified, stored, indexed, and retrieved for the purposes of reproducibility, credibility, and trustworthiness. Therefore, more advanced Object Identification mechanisms in the so-called *information clouds* are increasingly required. Table I compares centralized and decentralized PID systems for blockchain applications [32] and their potential for Transparency, Accountability, Robustness and other desired properties [33].

We propose integrating all three components (publication, code, datasets) in a supra descriptor that is added to a blockchain and stored with IPFS. We shall call it the *DOIschain*. That is to say, we have those three components coming from reproducible research that are associated with their own PIDs (DOI, SWHID, DataMed-ID), and we combine them into a single supra PID that integrates them and records them in a single blockchain transaction. It allows for reproducibility, trustworthiness, legitimacy, authenticity, and not censored robustness.

A. Generic Basic PID description for medical datasets and IPFS storage

The work by Brower and Narlock [34] draws attention to the special requirements for scientific data PIDs. Especially for face issues revealed by the COVID pandemic along with the possibility of using blockchain, it is possible to specify a specialized PID in medical datasets, given that there is a gap that makes it difficult for scientists to correctly associate PIDs of the final version of medical datasets used to obtain research results and present them with credibility and verifiability. In this case, an approach towards IPFS distributed storage guarantees the permanence, decentralization, redundancy, and robustness of these medical datasets and a PID in the blockchain that attests to its authenticity and legitimacy.

Let us give an example of a PID that could be used for medical datasets on IPFS, which is tailored to these medical needs and according to the data types handled. Table II presents the meta-descriptor for the basic PID for medical datasets, with the fields according to the possibilities found in the datasets commonly used in the medical field and according to the types [25] [24].

IV. Microservices Blockchain DOIschain prototype

We have designed an API to create and manage the proposed DOIschain PIDs.

A. Management of DOIschain

The prototype has an Ethereum private Blockchain consisting of three nodes representing authorities of the scientific community issuers of DOIschain (universities, publishers, laboratories) in a PoA configuration. A more high-level description of the API is given in Figure 1, showing the interactions with the IPFS and the Ethereum Private Blockchain. The use of IPFS prevents repeated uploading of the same data with its data-duplication-avoidance feature [33]. First, client parties (authors) upload their corresponding original data to IPFS. Then, their contributions saved in supra PID DOIchain are recorded in the blockchain by the authorities in the network. In

TABLE I
COMPARISON OF CENTRALIZED VS. DECENTRALIZED PIDS SYSTEMS WITH PROOF OF AUTHORITY (POA)

Aspect	Centralized (PID)	Decentralized PID
Control	Single authority controls the system	Distributed control among nodes
Scalability	Limited scalability	Potentially higher scalability
Trust	Relies on trust in the central authority	Trustless or trust minimized through consensus
Security	Single point of failure	Less susceptible to single points of failure
Efficiency	Central authority may improve efficiency	May have slower transaction throughput
Governance	Controlled by central entity	Community or consensus-driven governance
Transparency	Limited transparency	High transparency through public ledger
Resilience	Vulnerable to central authority failure	More resilient to individual node failures
Redundancy	Limited redundancy	High redundancy through multiple nodes
Robustness	Lower robustness due to central point of failure	Higher robustness due to distributed nature
Authenticity	High, due to central control	Ensured through consensus mechanisms
Legitimacy	Varies, depending on central authority's reputation	Based on decentralized consensus
Examples	ARK, DOI, SWHID	DID, CID, DOIschain

TABLE II
PROPOSED META DESCRIPTOR AND BASIC DATASET PERENNIAL PERSISTENT IDENTIFICATOR OF MEDICAL DATASETS IPFS STORED.

Field	Description	Example
ID	The unique identifier or accession number assigned to the dataset.	CID-QmSgtne642z5FTj7..
Dataset Name	The title of the medical dataset.	sample validation BraTS dataset
Description	A brief overview or summary of the dataset, including its purpose, content, and potential uses.	This dataset contains 4 test MRI images to test the execution of BraTSAPP
Source, Author	The origin or provider of the dataset, such as a research institution, government agency, or healthcare organization.	BraTS 2023 Challenge
Creation Date, Version	Control Version of the dataset.	2023
Format	The file format(s) in which the dataset is available, such as CSV, XML, JSON, etc.	.nii
Size	The size of the dataset in terms of storage space required, in bytes.	21474836480
Variables	The variables or features included in the dataset, along with their descriptions and data types.	T1, T2, FLAIR, T1ce
License	The licensing terms under which the dataset is distributed, indicating any restrictions or permissions for use.	Creative Commons CC BY-SA 4.0
MD5 Hash	The cryptographic hash value of the dataset file to ensure data integrity.	QmSgtne642z5FTj7..

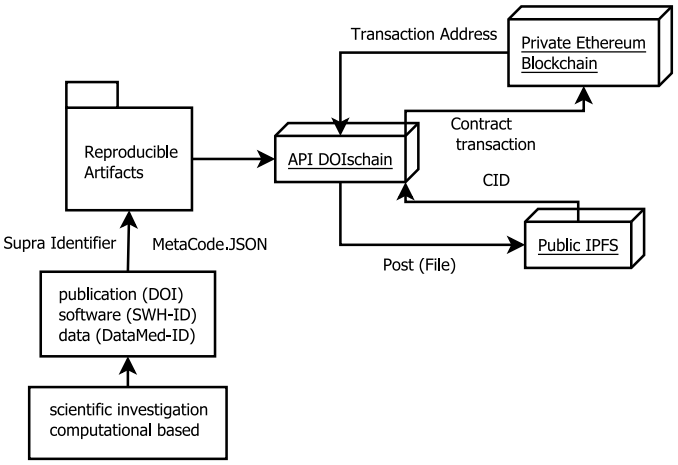

Fig. 1. API DOIschain high-level description. The computational products of the research (publication, code, datasets) are assigned a supra identifier recorded in a blockchain transaction.

this way, originality and transparency in the published research work are also guaranteed.

All transactions of issuing DOIschain are stored in the blockchain Proof of Work (PoA) so that traceability and versioning will remain persistent, immutable, and permanent for the long term, without intervention or decisions of a private or governmental nature. This handles a JSON request with the essential data from a front-end client authorized to issue DOIs described in Figure 2. We based our format on the existing CodeMeta generator tool by Software Heritage to incorporate the corresponding PID for the article (DOI), code (SWH-ID), and datasets IPFS PID in DOIschain.

B. Reproducible artifacts and metadata in IPFS

As described in Figure 1. All the required reproducible artifacts and the codemeta.json metadata are stored in the IPFS, which will be accessible through IPFS gateways[25] and persistently referenced by a hash CID v1. It must be recorded in the private blockchain ledger by one of the agreed authorities able to issue DOIs to be valid.

This guarantees that the artifacts will be persistently available and accessible over a long period of time regardless of the technological infrastructure or the decisions of the organization that issued the DOIs.

[25]https://ipfs.github.io/public-gateway-checker/

V. Use case: a BraTS medical application

We present a use case in medical field where we assign blockchain DOIschain to the reproducible artifacts of BraTS[26] International Brain Tumor Segmentation challenge scientific articles, that is, assigning PIDs to the results of the finished research product (scientific article, code, artifacts, data dataset) to be referenced, cited, and reproduced by the scientific community.

Let us start with a user story:

"As a radiologist, I wish to have updated and increasingly accurate neural network models available in emergency rooms to assist in interpreting brain magnetic resonance images. So, a diagnosis of the patient's severity is required very quickly"

For reproducibility and benchmarking MlCubes[27] is used. The special feature of this case is that MLcubes from MLcommons are used in the competition to containerize the model to make it reproducible and do benchmarking with the other contesters.

Additionally, the purpose was to create an end-to-end application for segmenting brain glioma called BraTSApp. In this case, the dataset is BraTS 2023 Glioma, and the neural network architecture is 3D-Unet.

The research results was presented in a publication for LAWCN2023 [35][28] and BrainLes2023[29].

The complete application, BraTSApp, with a graphical interface, was presented as an online reproducible and identifiable artifact example and product of the research in tumor segmentation.

The primary project artifacts and metadata issued from the medical research (codemeta.json [30] [2], plaintext source code, MLCube conf) are stored permanently in IPFS and assigned a persistent CID on the private Ethereum blockchain, which can be validated as a transaction in the blockchain.

Finally, as a result of the proof of concept, the entire computational research project developed for the BraTS competition in its final version for the scientific community has been assigned its perennial persistent identifier DOIschain global descriptor, with its respective publication article, code/model (MLCube) and dataset medical images (generic DataMed-ID) with persistent identifiers. The authenticity of the DOIschain is registered and validated in the private blockchain of the decentralized research entities with the authority to issue the identifier.

1) **DOIschain**[31]
   a) **Publication** [35][32]
   b) **Data**[33]

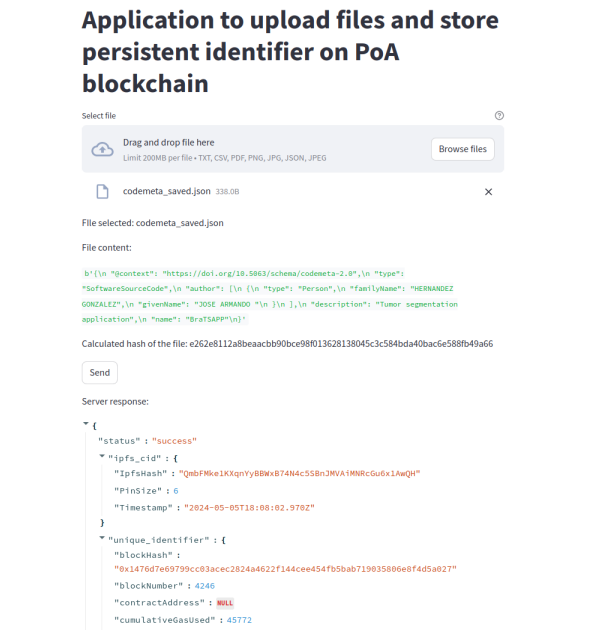

Fig. 2. Frontend API issuing DOIschain. Shows the process of issuing to DOIschain and the corresponding validation of transaction in Blockchain.

   i) **Generic DataMed-ID**: Table II
c) **Code**
   i) **BraTSAPP**[34]
   ii) **MLcube**[35]

In this way, we can contribute to the reproducibility, credibility, authenticity, FAIR compliance, consistency, comprehensiveness, and trustworthiness of the research results and claims.

VI. Conclusion

A generic basic but specialized perennial persistent identifier PID for medical data IPFS stored has been described, and management with a supra descriptor in the blockchain of all the elements resulting from scientific research based on computation is proposed.

As a use case, a medical application prototype has been built (within the framework of BraTS competition) with traceability of all produced artifacts by a supra descriptor of publication, data, and software, through persistent and perennial DOIschain supported by blockchain, which guarantees its long-term preservation, trustworthiness, attribution and contributes to its reproducibility.

References

[1] Stall, "Journal Production Guidance for Software and Data Citations," *Scientific Data*, vol. 10, no. 1, p. 656, Sep. 2023. [Online]. Available: https://www.nature.com/articles/s41597-023-02491-7

[26] https://www.synapse.org/Synapse:syn51156910/wiki/621282
[27] https://www.synapse.org/#!Synapse:syn51156910/wiki/622674
[28] https://link.springer.com/conference/lawcn
[29] https://link.springer.com/conference/iwb
[30] https://codemeta.github.io/codemeta-generator/
[31] https://white-historic-chinchilla-599.mypinata.cloud/ipfs/QmUBP3JvR5ombVPPQcBQZkBBPY9GzRtEgJBJZyQCx1RmuD
[32] https://white-historic-chinchilla-599.mypinata.cloud/ipfs/Qmezmd76kUg38qHmfb3xWXJyvJRGhmKJMKeGLfTMFjoPdj
[33] https://white-historic-chinchilla-599.mypinata.cloud/ipfs/QmSgtne642z5FTj71qHCeFoqUJPim3TxxRpBHcLV9JcaWV/
[34] https://white-historic-chinchilla-599.mypinata.cloud/ipfs/QmcEd9iZXx781RfeHKxpWcLQfiomMXukjEFpQoW6oKzo8B/
[35] https://white-historic-chinchilla-599.mypinata.cloud/ipfs/QmY8V23yiqABgrhnNFfwmT9WjAvjdnhiQ64HgFreoxiWFf/

[2] Research Data Alliance/FORCE11 Software Source Code Identification WG, A. Allen, A. Bandrowski, P. Chan, R. D. Cosmo, M. Fenner, L. Garcia, M. Gruenpeter, C. M. Jones, D. S. Katz, J. Kunze, M. Schubotz, and I. T. Todorov, "Use cases and identifier schemes for persistent software source code identification (V1.0)," 2020. [Online]. Available: https://zenodo.org/record/4312464#.X9BuzGgza70

[3] Rueda, "DataCite: Lessons Learned on Persistent Identifiers for Research Data," *International Journal of Digital Curation*, vol. 11, no. 2, pp. 39–47, Jul. 2017. [Online]. Available: http://ijdc.net/article/view/11.2.39

[4] Bönisch, "swMATH - a new information service for mathematical software," Jun. 2013, arXiv:1306.1036 [cs]. [Online]. Available: http://arxiv.org/abs/1306.1036

[5] K. Hinsen, "The Magic of Content-Addressable Storage," *Computing in Science & Engineering*, vol. 22, no. 3, pp. 113–119, May 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8887277/

[6] Hilse, *Implementing persistent identifiers: overview of concepts, guidelines and recommendations*. London: Consortium of European Research Libraries, 2006.

[7] Cosmo, "Identifiers for Digital Objects: the Case of Software Source Code Preservation," Sep. 2018, p. 1. [Online]. Available: https://hal.science/hal-01865790

[8] L. P, "A Universally Unique IDentifier (UUID) URN Namespace," RFC Editor, Tech. Rep. RFC4122, Jul. 2005. [Online]. Available: https://www.rfc-editor.org/info/rfc4122

[9] Confederation Of Open Access Repositories, "COAR Community Framework for Best Practices in Repositories." [object Object], Tech. Rep., Oct. 2020. [Online]. Available: https://zenodo.org/record/4110829

[10] D. Reed, M. Sporny, M. Sporny *et al.*, "Decentralized identifiers (dids) v1.0, core architecture, data model, and representations," https://www.w3.org/TR/did-core/, 2021.

[11] Schaufelbuhl, "EUREKA – A Minimal Operational Prototype of a Blockchain-based Rating and Publishing System," in *2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. Seoul, Korea (South): IEEE, May 2019, pp. 13–14. [Online]. Available: https://ieeexplore.ieee.org/document/8751445/

[12] Shinde, "Doc-Chain: A University Document Verification Blockchain," in *2022 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS)*. Pune, India: IEEE, Sep. 2022, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/9935950/

[13] Radha, "Verifiable Badging System for scientific data reproducibility," *Blockchain: Research and Applications*, vol. 2, no. 2, p. 100015, Jun. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2096720921000105

[14] Jha, "Certifier Dapp - Decentralized and Secured Certification System Using Blockchain," in *2023 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS)*. New Raipur, India: IEEE, Oct. 2023, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/10346509/

[15] J. Almeida and V. Amaral, "Towards trustworthy tracing responsibility of collaborative software engineering artefacts of student's software projects," in *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*. Los Alamitos, CA, USA: IEEE, Jun. 2022, pp. 151–160. [Online]. Available: https://ieeexplore.ieee.org/document/9842543/

[16] E. Bellini, "A blockchain based Trusted Persistent Identifier system for Big Data in Science," *Foundations of Computing and Decision Sciences*, vol. 44, no. 4, pp. 351–377, Dec. 2019. [Online]. Available: https://www.sciendo.com/article/10.2478/fcds-2019-0018

[17] Sicilia, "Decentralized Persistent Identifiers: a basic model for immutable handlers," *Procedia Computer Science*, vol. 146, pp. 123–130, 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1877050919300924

[18] A. Kumar and V. P. Kumar, "An Approach to Secure Decentralized Storage System Using Blockchain and Interplanetary File System," in *2023 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS)*. New Raipur, India: IEEE, Oct. 2023, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/10346339/

[19] P. A. Lobo and V. Sarasvathi, "Distributed File Storage Model using IPFS and Blockchain," in *2021 2nd Global Conference for Advancement in Technology (GCAT)*. Bangalore, India: IEEE, Oct. 2021, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/9587537/

[20] Sun, "Blockchain-Based Secure Storage and Access Scheme For Electronic Medical Records in IPFS," *IEEE Access*, vol. 8, pp. 59 389–59 401, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9045940/

[21] Gao, "Research on Medical Image Data Sharing Traceability System Based on Blockchain," in *2023 5th International Conference on Decision Science & Management (ICDSM)*. Changsha, China: IEEE, Mar. 2023, pp. 113–116. [Online]. Available: https://ieeexplore.ieee.org/document/10314107/

[22] Lin, "Blockchain-based Secure Storage System for Medical Image Data," in *2023 IEEE 3rd International Conference on Electronic Communications, Internet of Things and Big Data (ICEIB)*. Taichung, Taiwan: IEEE, Apr. 2023, pp. 158–163. [Online]. Available: https://ieeexplore.ieee.org/document/10170051/

[23] R. F, "Medical subject headings," *Bull Med Libr Assoc.*, pp. 114–6., 1963.

[24] Chen, "LitCovid: an open database of COVID-19 literature," *Nucleic Acids Research*, vol. 49, no. D1, pp. D1534–D1540, Jan. 2021. [Online]. Available: https://academic.oup.com/nar/article/49/D1/D1534/5964074

[25] Shuja, "COVID-19 open source data sets: a comprehensive survey," *Applied Intelligence*, vol. 51, no. 3, pp. 1296–1325, Mar. 2021. [Online]. Available: https://link.springer.com/10.1007/s10489-020-01862-6

[26] Sansone, "DATS, the data tag suite to enable discoverability of datasets," *Scientific Data*, vol. 4, no. 1, p. 170059, Jun. 2017. [Online]. Available: https://www.nature.com/articles/sdata201759

[27] Ohno-Machado, "Finding useful data across multiple biomedical data repositories using DataMed," *Nature Genetics*, vol. 49, no. 6, pp. 816–819, Jun. 2017. [Online]. Available: https://www.nature.com/articles/ng.3864

[28] Wimalaratne, "Uniform resolution of compact identifiers for biomedical data," *Scientific Data*, vol. 5, no. 1, p. 180029, May 2018. [Online]. Available: https://www.nature.com/articles/sdata201829

[29] c.-i. T. G. o. D. C. S. a. Practices, "Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data," *Data Science Journal*, vol. 12, no. 0, Sep. 2013. [Online]. Available: https://datascience.codata.org/articles/10.2481/dsj.OSOM13-043

[30] Data Citation Synthesis Group, "Joint Declaration of Data Citation Principles," [object Object], Tech. Rep., 2014. [Online]. Available: https://www.force11.org/group/joint-declaration-data-citation-principles-final

[31] Xia, "Challenges in decentralized name management: the case of ENS," in *Proceedings of the 22nd ACM Internet Measurement Conference*. Nice France: ACM, Oct. 2022, pp. 65–82. [Online]. Available: https://dl.acm.org/doi/10.1145/3517745.3561469

[32] V. Buterin, "A next generation smart contract & decentralized application platform," 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:19568665

[33] Zhang, "Exploiting Blockchain to Make AI Trustworthy: A Software Development Lifecycle View," *ACM Computing Surveys*, vol. 56, no. 7, pp. 1–31, Jul. 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3614424

[34] Brower, "Persistent Identifiers and Research Data," in *2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. Santa Fe, NM, USA: IEEE, Jun. 2023, pp. 269–270. [Online]. Available: https://ieeexplore.ieee.org/document/10265951/

[35] J. A. Hernández, "Data Augmentation by Adaptative Targeted Zoom for MRI Brain Tumor Segmentation," in *Computational Neuroscience*, R. Salas, Ed. Cham: Springer Nature Switzerland, 2024, vol. 2108, pp. 14–24. [Online]. Available: https://link.springer.com/10.1007/978-3-031-63848-0_2