dPIDs - the Emerging Persistent Identification Technology for FAIR and the Digital Era

Andrey Vukolov
Scientific Computing Group
Elettra Sincrotrone Trieste
Trieste, Italy
0000-0001-6967-3236

Erik van Winkle
GOFAIR Foundation Fellow
DeSci Labs AG
Bäch, Switzerland
0000-0002-7567-0311

Elizaveta Zhdanova
Faculty of Fine Arts
Valencia Polytechnic University
Valencia, Spain
0009-0005-3701-0266

Georgios Kourousias
Scientific Computing Group
Elettra Sincrotrone Trieste
Trieste, Italy
0000-0002-1243-7168

***Abstract*—This paper presents a comprehensive exploration of an emerging technology focused on the persistent identification and sharing of data and metadata, termed "decentralised Persistent Identifier" (dPID). Utilizing the InterPlanetary File System (IPFS) network, dPID provides reproducible persistent identifiers, ensuring that data can be reliably stored and accessed over time. It is designed to incorporate a distributed ledger, Public Key Infrastructure (PKI), decentralized linking, and lookup databases. It also aims to support version control and provenance tracking based on reproducibility. The distributed approach highlights the potential of dPID to align with the principles and guidelines of Findable, Accessible, Interoperable, and Reusable (FAIR) data, positioning it as a valuable component within the open data ecosystem, mitigating in an efficient way the problems, such as link rot and content drift. The proposed technology provides a highly scalable persistent identification and versioning system for shared data that reduces dependencies from social contracts and institution-driven systems. The paper also demonstrates its usability for modern social-oriented identification systems, proposing a use case study for the art industry.**

***Index Terms*—PID, Persistent Identifier, IPFS, Decentralized, dPID, FAIR, Blockchain, Provenance Tracking, Reproducibility**

I. Introduction

Persistent Identifiers (PIDs) are not only for documents and data. They can also be used to reference other entities or agents, including people contributing to the research, software, research organizations, physical objects such as samples and instruments, and even abstract concepts such as terms in a controlled vocabulary. As a general rule, whenever something needs to be referenced in a reliable and lasting manner, a persistent identifier should be used [1].

The modern world is encountering unprecedented challenges in data handling. Growing amounts of data in the realm of academia, including Internet of Things (IoT) sensing, multi-disciplinary works and Artificial Intelligence (AI) form a collective ocean of knowledge, impassible without the collaboration of humans and machines. One point of advocacy for the necessary human-computer symbiosis can be found in the Findable, Accessible, Interoperable, and Reusable (FAIR) principles and subsequent literature around machine-actionability [2], [3]. Policy based recognition in support of these principles has been released over the past 5 years from organizations such as Office of Scientific and Technological Policies (OSTP), European Commission (EC), and National Institutes of Health (NIH) [4]–[6].

The first of the 15 FAIR Principles, Principle F1, states that "Data and metadata must have globally unique, persistent, and resolvable identifiers" [2]. In the upcoming world of FAIR Science, the scientific record strives for the permanence of information online. While permanence is impossible, PIDs and underlying technological infrastructures are essential components of continuous navigation and data retrieval. Existing PID providers such as Digital Object Identifier (DOI), mitigate the problems of persistence using social and technical mechanisms to guarantee consistent mapping and resolution of a given ID to its target resource. This goal of persistent identification and resolution has a standard set of problems.

Ensuring the scalability of social and technical mechanisms in persistent identification technology can be a challenge. Industry 4.0, IoT and rapidly growing open-source development [7] drastically increase the amount of data that should be identified and linked in a strictly deterministic manner. The reusability of metadata is required for the efficiency of data access, meaning that the persistence of metadata and its subsequent linkage to data becomes essential too. Social trust as a mechanism to combat the replication crisis requires added persistence requirements. The high diversity of data storage and transportation techniques, protocols, networks, etc. also reveals new data treatment circumstances where every file should be identified, checked for integrity, versioned, and shared over many possible networks with different address resolution mechanisms. The average scientific project or experiment outcome may contain hundreds of thousands of files (research artifacts), including code, data, reviews, sensemaking statements, etc., both created by the people and captured automatically. Global endeavors in a single scientific domain can produce trillions of individual files of their own accord. The pursuit of permanence for scientific knowledge starts with a problem of scaling persistent identification infrastructure.

Additional challenges surrounding current PID infrastructure include but are not limited to the sovereignty and access control over data [8], the necessity of machine-actionable provenance, client-side reproducibility of existing PIDs due to their generation algorithms being enclosed within the registrar's infrastructure and an exponentially growing amount of

 ISBN: 978-1-68558-144-2

resolution requests due to the formation of PID graphs [9].

With trillions of PIDs likely needed by 2030, the probability of system failure in current federated but centralized architectures lead to system fragility and a potential loss of valuable knowledge [10].

To stay relevant and reliable under the extreme circumstances of science in the digital age, the PID system's intended usage in the modern FAIR-compliant data-driven Internet needs to implement a reproducible, fully automated, scalable, globally unique identifier system, with minimal reliance on social contracts, leading to fewer centralized points of failure and human interactions. In the following sections, we will explore the range of currently available PID systems, delving into their underlying technologies and unique characteristics. Subsequently, we will present a technical proposal for an emerging decentralized PID system, including a concise analysis and discussion on the prospects of its implementation.

Additionally, this paper proposes a potential application of the dPID technology in the realm of fine arts and photography. This use case focuses on hybrid distribution models that encompass both hardcopy and digital formats, addressing the unique challenges of managing, distributing, and verifying artworks across diverse mediums. By leveraging dPID, stakeholders in the fine arts sector could benefit from enhanced provenance tracking, more secure distribution channels, and improved accessibility to art, all while adhering to the FAIR data principles.

The rest of the paper is structured as follows. In Section 2, we explain the current status of the PID implementation landscape, holding European Open Science Cloud Photon and Neutron Data Services (ExPaNDS) project outcomes as a basis. In the first part of Section 3, we give a technical proposal based on one of the possible implementations of a decentralised approach to persistent identification. In the second part of this section, we leverage the technology by giving a technical outline of freshly developed open-source software. Then in the Conclusion section, we summarize the problems we have indicated solved in the paper. This section is also extended with a use case proposal showing the possibilities of dPID in the area of persistent identification and distribution of the artworks.

II. Current PID Implementations: Landscape and Centralization Status

This section is partially based on works [1] and [11]. It explores the implementation details of the most known and popular PID systems in the context of centralization. The mentioned PID systems: DOI, Open Researcher and Contributor Identifier (ORCID), Handle, Research Organization Registry (ROR), etc., implemented worldwide fall between *centralized* and *federated* architectures [12] implementing the centralised resolution model presented in Figure 1. The decentralised architectures are now only emerging, so none of the existing systems could be named decentralised.

The model in Figure 1 implements a low number of high-risk singular points of failure, (as it is based on the entity called **Global PID Registry**), despite the existence of both primary and secondary authorities. From the client's point of view, they should be called **centralized**. Every existing point of failure here may lead to the failure of the entire facility or domain. If the PID is not reproducible and the generation schema is closed, the centralised PID governance authorities are the only provenance holders of the underlying record, metadata, and addressed data [13]. This self-assignment trap magnifies when record tracking lags behind the growing number of maintained records. Issues such as link rot, content drift, artifact fragmentation and inconsistent resolution [14] in these systems can be traced back to human interactions, leading to lacking persistence because they do not provide lookup table redundancy [15], [16], reproducibility, and proper caching. However, while the extreme circumstances of the modern data-driven world do not affect all domains, centralized authorities such as person-oriented ORCID and organization-oriented ROR have proven trustworthy over time, at scales of 1-10M of PIDs. For their social-driven data flows and scalability cases, they are simple, efficient, and can be considered fit-to-purpose.

As a reply towards the challenges arising, the different flavours of **federated** approach have appeared. It combines the existing conservative model of centralized governance authority with a federated social-driven network of formally independent PID registrars. Each registrar manages the underlying system of centralized prefix-based lookup tables with a federated network of resolvers based on Hypertext Transfer Protocol (HTTP) redirection. They act from the governance point of view, as independent registries with independent databases replicated at the discretion of the centralized authority. Each registrar maintains its own provenance authority and ensures the persistence of associated PIDs alone. To control the resources distributed across the federation of infrastructure providers in maintaining persistent registry infrastructure, an external council or legal entity is still needed, leading to a non-profit governance ogranization overseeing the system. Through this structure, the federated PID infrastructure aims to preserve the administrative efficiency of the centralized approach while reaching redundancy and distributed curation of decentralized approaches. The most popular PID systems in the world, DOI and Handle.net, use the federated approach. It allows expansion of the computational abilities of the system, primarily in the aspect of data replication.

To obtain interoperability and cross-resolution, especially in federated architectures, the third-party aggregation approach, in practice, is the only way. The UNIRESOLVER initiative [17] is the most known solution for implementing it. It utilizes Worldwide Web Consortium (W3C) Decentralized Identifiers (DID) specification [18]–[20] to perform a bidirectional lookup but without internal implementation of metadata versioning and reproducibility tracking, so the records resolved via UNIRESOLVER could not be considered truly immutable and the associated resolution cannot be considered deterministic on the periods on which the PIDs should persist. However, UNIRESOLVER demonstrates the possibilities of cross-resolution and request flow balancing in the federated

 ISBN: 978-1-68558-144-2

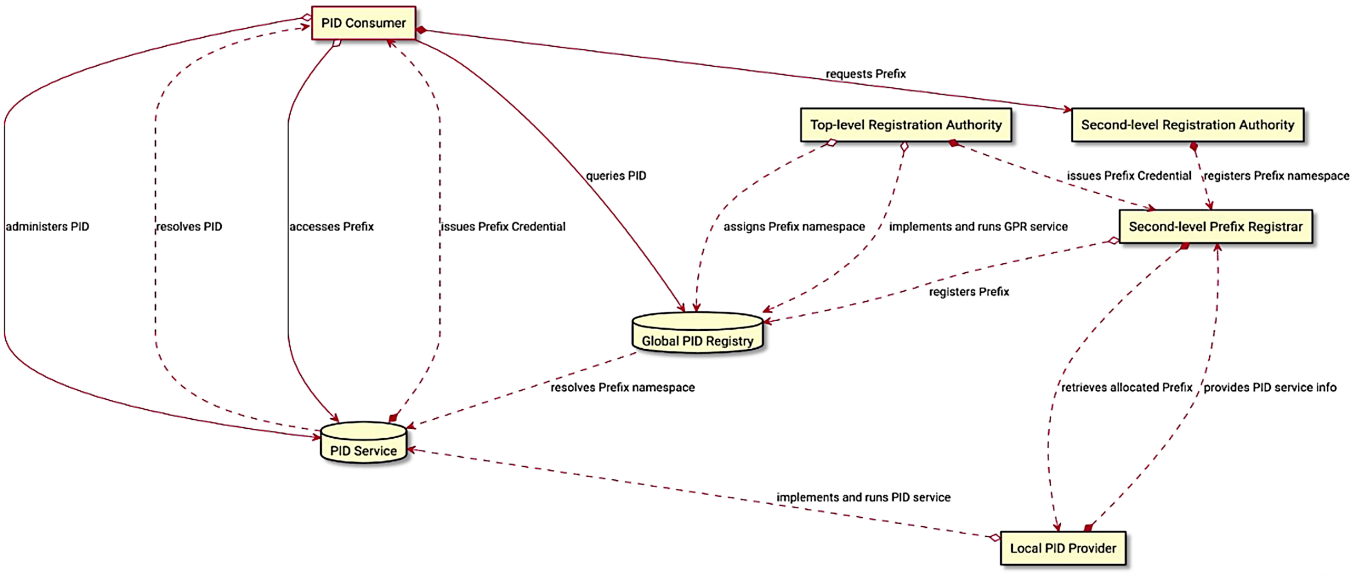

Fig. 1. Traditional PID administration and implementation schema.

architecture without rebuilding the existing provenance chains.

In Table I, a brief list of currently implemented worldwide PID providers is presented, with their underlying technologies and/or PID type specifications and dedicated entity types. The table is taken from [1] with minor additions. In Table I:

- DAG — Directed Acyclic Graph.
- IGSN — International Generalized Sample Number.
- SWHID — Software Heritage Identifier.
- RAiD — Research Activity Identifier.

As it is easy to see from the entries presented in Table I, only one of the PID providers implemented worldwide uses a PID with client-side reproducibility — SWHID. Thus, it should be considered the only one that has already implemented the provable immutability and inline versioning, but under the centralized governance. The federated DOI system holds leadership in dissemination; it is used as the underlying provider by most listed PID providers.

III. dPID: Adding Decentralization to Data Identification Pipelines

The core idea behind decentralized data identification technologies is that the client should be asking a fundamentally different question when resolving the metadata from the PID. Instead of asking a single point: ”What is the content stored at this location?”, they should be asking a network-based decentralized swarm: “Can you tell me how to find the content with this hash?”. This approach is one of the natural continuations of an idea of the PID Graph [9], proposing the technical mechanisms as the primary mode of guaranteeing persistence, replacing the social mechanisms used in current systems. In a decentralized PID system, **the social persistence of a PID is as strong as the technical prevalence of the network nodes, providing self-describing addressing, resolution, and consistent data**. One of the possible workflow schemas implementing the decentralized approach is presented in Figure 2.

As it could be easily deducted from the block names in Figure 2, most of the elements of the PID system indicated are the distributed ledgers and databases where the data are not owned or stored on the side of the single participant or institution: the lingering in this schema requires a mathematically proven integrity of the internal metadata stored in the system. Also, the presented schema should not be considered as a reference: the decentralised approach makes the implementation schema fuzzy and yields its structure in favour of the workflow.

The ideal implementation of the PID system necessitates uniformity in backend technology and storage models across all resolvers and PID generators. In such a setup, users would need to install a specific software, possibly on a local machine, and register a public key. This process would grant them access to a resolution endpoint, equip them with a PID generator that supports namespace propagation right from the start, and provide a viewer for the provenance chain. It demands data pipeline technologies that can handle storage, identification, and delivery in a manner that is as type-agnostic and mathematically secure as possible. Fortunately, there exists a category of technologies that fulfil these requirements. Broadly, these can be characterized as decentralized redundant storage systems, with BitTorrent being one of the most well-known examples. Such technologies rely on hash-based, mathematically-driven content addressing, and utilize Distributed Hash Tables (DHT) for the delivery of binary data objects. DHT, in particular, incorporates identification directly within its addressing stack, making these technologies an apparent choice for the described purposes [18].

IPFS, the decentralized storage network [21] stands out by allowing data to be stored under a single identifier that is immutable and persistent, based on its binary representation.

 ISBN: 978-1-68558-144-2

TABLE I
INTERNATIONAL PID PROVIDERS, THEIR UNDERLYING TECHNOLOGIES AND ENTITY TYPES.

Provider	Technology	Entities	Centralization
DataCite	DOI	General purpose	Federated
Crossref	DOI	Publications, funders	Federated
ePIC	Handle	Metadata in all plaintext schemas	Federated
IGSN	Handle	Experimental samples	Federated
ORCID	Bespoke (custom)	Persons	Centralized
FigShare PID	DOI	Research artifacts	Federated
Zenodo	DOI	Publications, digital research artifacts	Federated
EUDAT B2SHARE	Handle, DOI	Datasets, digital research artifacts	Federated
FAIRshare	DOI	Datasets, policies, standards	Federated
SWHID	SHA1-based Merkle DAG	Software artifacts, versioned source code	Federated, decentralized
ROR	Bespoke (custom)	Research institutions	Centralized
RAiD	Handle (custom, prefixed)	Funders, organizations, persons, instruments, datasets	Centralized

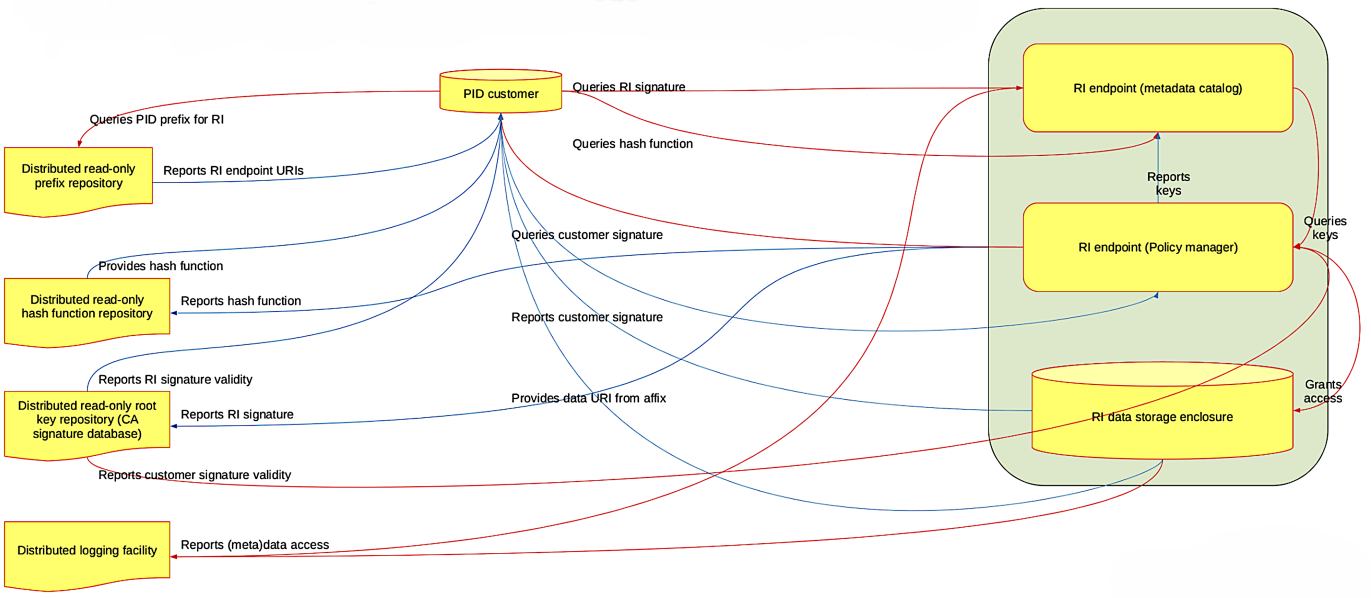

Fig. 2. One of the possible decentralised PID workflow implementation schemas.

This system employs Content Identifiers (CIDs), which are derived from the outputs of unidirectional cryptographic hash functions [22]. These IDs incorporate a prefix-based structure, with the specification detailing how they should be resolved, ensuring a comprehensive description within the ID's structure itself. CIDs can be seen as production-ready PIDs that are resistant to content drift and link rot.

The IPFS storage model is designed to support a cache-on-read retrieval approach, where data, along with its metadata, is recorded into DAGs using various storage tree builders. This ensures that data and metadata are automatically cached, with the option for explicit declaration of caching for specific data on a given network node. The persistence levels for data stored in IPFS, as determined by the CID's state declared in the DAG, are outlined in Table II. In this context, a "pinned" state indicates that a specific CID has been explicitly marked for caching on at least one node within the IPFS network. The term "Accessible" means that the requested data is available for download, and "Findable" means that the data will become available and propagating through the DAG at the moment when at least one node knowing the given CID has appeared in the network.

TABLE II
POSSIBLE AVAILABILITY STATES OF THE DATA ENTRY IN THE IPFS NETWORK.

IPFS DAG *Local data*	Pinned	Wanted	Available	Discarded
Stored	Persistent	Accessible	Persistent	Accessible
Requested	Persistent	Accessible	Persistent	Findable
Idle	Persistent	Accessible	Accessible	Accessible
Discarded	Findable	Findable	Findable	Unavailable

IPFS also incorporates the InterPlanetary Linked Data (IPLD) concept, utilizing a hash-based addressing model that facilitates the construction of decentralized storage trees [23]. This model enables a persistent linkage between stored items

and their versions, ensuring that each piece of data can be uniquely identified and accessed over time. Given the features that have already been implemented, IPFS emerges as a suitable choice for implementing the PID system, thanks to its robust framework for storing and linking data in a decentralized manner.

A. dPID: Technical Proposal

To establish a production-grade PID system, integrating IPFS with additional components that facilitate persistent linkage, version control, and contribution tracking is essential [24]. This paper introduces dPID, a decentralized data identification and sharing system designed to enhance the management and curation of research artifacts and FAIR Data Objects. dPID represents a pilot project aimed at offering a reliable solution for persistent storage and data curation [25].

At its core, dPID leverages a sophisticated integration of the IPFS and IPLD technologies alongside the Sidetree Protocol [26]. Sidetree acts as a protocol and Application Programming Interface (API) layer that can operate atop any data addressing system, enabling users to generate lookup databases and customize identities secured by PKI. Utilizing Sidetree, dPID adopts the standardized DID schema standardized by W3C, which facilitates the creation of universally unique identifiers that are resilient and verifiable [20].

dPID provides access to data that is both machine-actionable and human-readable, featuring a web interface built upon the JavaScript Object Notation Linked Data (JSON-LD) specification [27] and supported by an open-source API. The software suite responsible for resolving and minting dPIDs serves as a uniform kernel for every installation, ensuring that the system is completely open-source and operable on dedicated systems as-is. This setup not only enables efficient minting and resolution processes but also guarantees a public resolver functionally equivalent to any resolver within the decentralized network.

Enhancing the foundational persistence offered by IPFS, dPID introduces features like high throughput, strong consistency across the network, decentralized indexing, user-friendly URLs, and the incorporation of a Turing-complete blockchain. This blockchain component autonomously records the root CID of an IPLD data structure, ensuring metadata redundancy and immutability. Consequently, dPID promises reliable persistence for stored FAIR data objects and research artifacts, making it a comprehensive solution for decentralized data management and sharing. dPIDs provide:

- Verifiable ownership with ORCID-based person identification and incremental contribution record.
- Open network participation and metadata redundancy through peer-to-peer nature of IPFS.
- Compliance with FAIR principles via FAIR Data Object specification compatibility.
- “Vendor lock-in” removed in the context of data due to the removal of the singular provenance holder of the scientific record.
- Data integrity persistence with DHT, as it was described above.

B. dPID Nodes: Brief Technical Outline

dPID Nodes is an open source [28] software suite written in TypeScript and published under MIT license [29]. It acts over IPFS HTTP API that should be provided by an API server. The API server currently used is Kubo - the open-source reference implementation of the IPFS node that runs in the background provides CID resolution and retrieves the data from the decentralised network. dPID considers the identified data as a collection of versioned IPLD entries following Research Object Crate (RO-CRATE) specification [30]. The simplified example of internal linkage is presented in Figure 3. From the perspective of the end user, these systems can be likened to folders that store research artifacts in a format-agnostic manner, with authorship and provenance details facilitated through integration with ORCID. The storage entries are catalogued in a distributed key-value store, each uniquely accessible via a dedicated persistent identifier that leverages underlying IPFS CIDs for addressing. The links between CIDs in Figure 3 illustrate the internal linked data structure that implements versioning and the history of changes. According to IPFS specification, every CID included in this schema is immutable, so the dPID metadata actually formulates a versioned repository for every FAIR Data Object it addresses.

As outlined in the documentation [25], [31], dPID ensures deterministic resolution of PIDs to the internal CIDs of IPFS and their associated content through a DAG. This process allows the content to be immediately cached on the local database of the node it is accessed from. Building upon the IPFS framework, dPID Nodes additionally employ Ceramic [32], a decentralized event streaming protocol, to create a graph-based distributed lookup database. This integration facilitates advanced data addressing using a combination of CID and DID.

When dPID Nodes installation is queried through an HTTP API, it returns JSON-LD object [27]. These objects can then be resolved by the web frontend, presenting the data in a human-readable format. This mechanism ensures that end users can easily access and interpret the stored research artifacts, benefiting from a seamless integration of decentralized storage technologies and modern web standards for data representation and access. The system in its current state should be considered as a pre-beta pilot version, work in progress on the deployment process and features list.

C. Challenges on the Security

dPID is designed to be secure. Security is among the top priorities and there is high confidence in the current implementation due to the secure nature of its individual components. Openness and strong cryptography with options for transparent upgradability are already in place considering an openly shared source code. It is beyond the scope of this paper to analyse in depth the strengths and weaknesses of the technological choices in the scope of security. Nevertheless, it hints that they

 ISBN: 978-1-68558-144-2

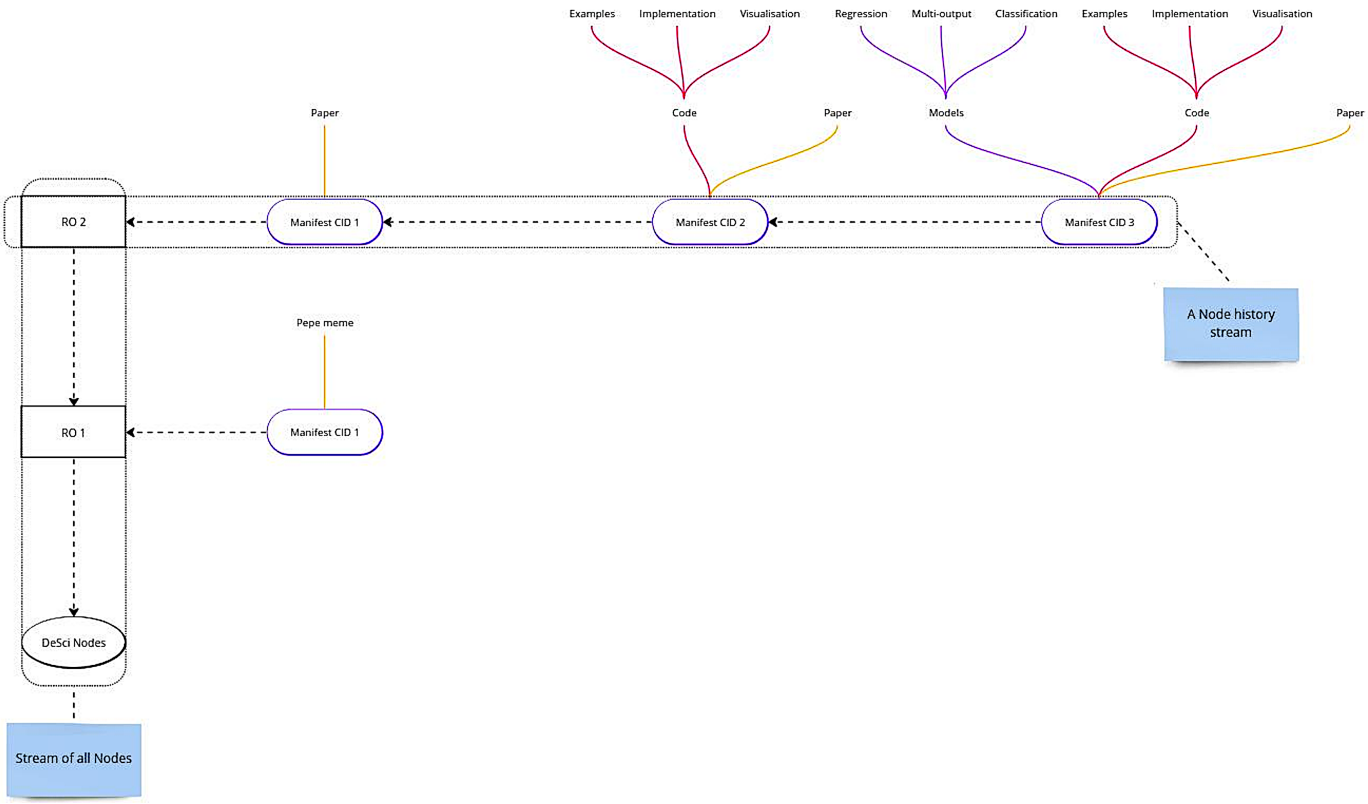

Fig. 3. Simplified internal bitstream storage diagram in dPID.

are taken into consideration. Those choices of technologies also aim at feasible disaster recovery and the robustness of the system. A recent incident demonstrated the readiness of the system and here it is briefly mentioned. Elettra Sincrotrone Trieste, like similar large facilities, had a security incident of unknown nature but with limited and contained damages. An isolated series of ransomware attacks required swift actions and temporary restriction of experimental and non-essential services to the operation of the facility. Such a service was that of the dPID gateway which albeit not compromised, it had to get offline. The underlying technology of dPID allowed for the rapid migration to a new server, outside the facility and hosted in a different country. The migration was rapid and easy (around 2 days including overrun and migration of the underlying CIDs and IPLD tracks), allowing for the continuation of the tests. Without focusing on the individual components that allowed this, the takeaway message is the decentralised nature of dPID which is in contrast to all other established, and especially, the centralised PID solutions.

IV. Roadmap for Further Development

The dPID initiative for the future strives to:

- Provide a set of convenience libraries with a comprehensive API and viable examples letting software developers adopt the technology.
- Propose a deployment pipeline feasible for different use cases (OS packages, bundles, automated source builders, Docker scripts, etc.).
- Develop comprehensive documentation for end users, publishers and administrators letting them adopt and use the technology in the most configurable and flexible way.
- Propose social-oriented mechanisms such as social attestation, conflict moderation, and access control.
- Formal validation of the security and robustness models of dPID.
- Explore and adopt multiple data-oriented use cases such as migration, storage sharing, and consortium mechanisms.

V. Conclusion and Use Case Proposal

A description of the dPID system and the problems it solves is provided above. The Nodes web interface is openly published and used to distribute research artifacts published with open licenses. The experimental resolution endpoint lives on the website [33]. For the installation and usage complexity, there are only qualitative estimations that exist now. However, because every participant of the dPID network should share completely the same software and all elements of the system, including the PID resolution point based on the website, the deployment complexity of the system should be estimated as approximately equal to the complexity of deployment

 ISBN: 978-1-68558-144-2

and debugging of the standard organizational website in the cloud. The deployment pipeline uses highly standardised and documented solutions, such as Docker, so the process can be effectively controlled from the developers' side, and volunteer support through Github is also available.

The research introduces a compelling use case for dPID as a foundational infrastructure for the identification and hybrid (hardcopy + digital) distribution of artworks, incorporating provenance tracking and the engagement of social institutions. This initiative aims to facilitate the management of the provenance chain for photographic artworks, among others, that are distributed in hard copies, as Non-Fungible Tokens (NFTs), and in digital formats. The system is designed to link the author's personal PID with the authorized digital copy of an artwork and the CIDs of copies authorized for distribution, accompanied by provenance chain documents validated by social institutions.

This model is particularly suited for electronic use, incorporating linked data that includes technical details to identify unauthorized digital distribution and support implicitly legal promotion mechanisms, such as search engines. Also, it defines the basic identification procedures and unified initial provenance for the hardcopy distribution of the digital artworks, with an option to extend the practices to the material artworks. Currently, the project is in the stage of defining its workflow model, with the metadata model already established. It is being explored as a potential application for dPID deployment, aiming for social recognition and validation.

Through this approach, dPID could offer a robust solution for artists and institutions to securely manage and distribute artworks. By integrating digital and hardcopy formats with a comprehensive provenance chain, the project seeks to enhance the trust and verifiability of artworks' distribution and ownership, leveraging decentralized technologies for greater transparency and security in the art world.

Due to the integrated attestation mechanisms, for external users, outcomes of the dPID project have approximately the same place as DOIs. Especially for the artworks, dPID extends the usual application area of the authenticity certificates, making them recognisable worldwide using any accessible resolution point held by any dPID network participant. From the author's point of view dPID has an outcome as a versioned, reproducible metadata handler for his authored data (digital photographs, articles, research objects, etc.). Acting locally, it simplifies minting and extends the functionality of systems like DOI, with out-of-the-box immutability.

ACKNOWLEDGMENT

Authors acknowledge **Linda Simeone** and **Andrey Kurin** for their contribution in the development of the artwork identification use case, especially in the social and legal areas.

REFERENCES

[1] V. Bunakov, R. Krahl, B. Matthews, N. Vizcaino, and A. Vukolov, "Advanced infrastructure for PIDs in Photon and Neutron RIs," Mar. 2022.

[2] M. D. Wilkinson *et al.*, "The FAIR guiding principles for scientific data management and stewardship," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[3] T. Bozada *et al.*, "Sysrev: A FAIR platform for data curation and systematic evidence review," *Frontiers in Artificial Intelligence*, vol. 4, 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frai.2021.685298

[4] B. M. Kuehn, "NIH Data Sharing," *JAMA*, vol. 298, no. 20, pp. 2361–2361, 11 2007. [Online]. Available: https://doi.org/10.1001/jama.298.20.2361-a

[5] R. David *et al.*, "Be sustainable: EOSC life recommendations for implementation of FAIR principles in life science data handling," *The EMBO Journal*, vol. 42, no. 23, p. e115008, 2023. [Online]. Available: https://www.embopress.org/doi/abs/10.15252/embj.2023115008

[6] H. Moulaison-Sandy, "The Nelson Memo and US Federal Funder Requirements for Public Access: Implications for Technical Services Librarians," *Technical Services Quarterly*, vol. 40, no. 4, pp. 290–297, 2023. [Online]. Available: https://doi.org/10.1080/07317131.2023.2271278

[7] S. Plaga *et al.*, "Securing future decentralised industrial iot infrastructures: Challenges and free open source solutions," *Future Generation Computer Systems*, vol. 93, pp. 596–608, 2019.

[8] P. N. Mahalle, G. Shinde, and P. M. Shafi, *Rethinking Decentralised Identifiers and Verifiable Credentials for the Internet of Things*. Cham: Springer International Publishing, 2020, pp. 361–374.

[9] H. Cousijn *et al.*, "Connected research: The potential of the pid graph," *Patterns*, vol. 2, no. 1, 2021.

[10] A. Vukolov, "Openly reproducible Persistent Identifiers (PIDs) as a factor of FAIRness in data sharing practices," Jun. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.4980522

[11] P. de Castro, U. Herb, L. Rothfritz, and J. Schöpfel, "Some reflections on the current PID landscape – with an emphasis on risks and trust issues," *Procedia Computer Science*, vol. 211, pp. 28–35, 2022, 15th International Conference on Current Research Information Systems. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050922016386

[12] J. Brown, "PID federation scoping study: final report," Sep. 2020.

[13] N. Soler *et al.*, "Final recommendations for FAIR Photon and Neutron Data Management," Jul. 2022.

[14] M. Klein and L. Balakireva, "On the persistence of persistent identifiers of the scholarly web," in *Digital Libraries for Open Knowledge*, M. Hall, T. Merčun, T. Risse, and F. Duchateau, Eds. Cham: Springer International Publishing, 2020, pp. 102–115.

[15] J. Klump *et al.*, "Towards globally unique identification of physical samples: Governance and technical implementation of the IGSN global sample number," *Data Science Journal*, vol. 20, no. 1, pp. 1–16, 2021.

[16] J. Kunze, "Towards electronic persistence using ARK identifiers," UC Office of the President, 2003. [Online]. Available: https://escholarship.org/uc/item/3bg2w3vs

[17] UNIRESOLVER, "Code Repository," Github, access date: 23.01.2024. [Online]. Available: https://github.com/decentralized-identity/universal-resolver

[18] M.-A. Sicilia, E. García-Barriocanal, S. Sánchez-Alonso, and J.-J. Cuadrado, "Decentralized persistent identifiers: a basic model for immutable handlers," *Procedia Computer Science*, vol. 146, pp. 123–130, 2019, 14th International Conference on Current Research Information Systems, CRIS2018, FAIRness of Research Information.

[19] DIDs, "Decentralized Identifiers v1.0. Core architecture, data model, and representations. W3C Recommendation 19 July 2022," Website, access date: 22.01.2024. [Online]. Available: https://w3c.github.io/did-core/

[20] N. Bach, "Dezentrale identifikatoren (dids): Die nächste pid-evolution: selbstsouverän, datenschutzfreundlich, dezentral," *o-bib. Das offene Bibliotheksjournal/Herausgeber VDB*, vol. 8, no. 4, pp. 1–20, 2021. [Online]. Available: https://doi.org/10.5282/o-bib/5755

[21] IPFS, "Official Documentation," Website, access date: 21.01.2024. [Online]. Available: https://docs.ipfs.tech/

[22] CID, "Specification," Website, access date: 21.01.2024. [Online]. Available: https://github.com/multiformats/cid

[23] IPLD, "Official Documentation," Website, access date: 23.01.2024. [Online]. Available: https://ipld.io/docs/

[24] A. Niehues *et al.*, "A multi-omics data analysis workflow packaged as a FAIR Digital Object," *GigaScience*, vol. 13, p. giad115, 01 2024. [Online]. Available: https://doi.org/10.1093/gigascience/giad115

 ISBN: 978-1-68558-144-2

[25] dPID, “Codex,” Website, access date: 23.01.2024. [Online]. Available: https://codex.desci.com/desci-codex/design-goals
[26] Sidetree, “Protocol documentation,” Website, access date: 22.01.2024. [Online]. Available: https://identity.foundation/sidetree/spec/
[27] JSON-LD, “Documentation,” Website, access date: 21.01.2024. [Online]. Available: https://json-ld.org/learn.html
[28] dPID, “Nodes,” Code repository, access date: 23.01.2024. [Online]. Available: https://github.com/desci-labs/nodes
[29] “MIT License,” Website, access date: 23.01.2024. [Online]. Available: https://opensource.org/license/mit/
[30] RO-CRATE, “Research Object Crate,” Official Documentation, access date: 21.01.2024. [Online]. Available: https://w3id.org/ro/crate
[31] dPID, “Documentation,” Website, access date: 22.01.2024. [Online]. Available: https://docs.desci.com/
[32] Ceramic, “Protocol Documentation,” Website, access date: 23.01.2024. [Online]. Available: https://developers.ceramic.network/docs/protocol/js-ceramic/overview
[33] “dPID Nodes Web Interface,” Website, access date: 23.01.2024. [Online]. Available: https://nodes.desci.com

 ISBN: 978-1-68558-144-2