

A Cost-Aware Blockchain–IPFS Architecture for Provenance of Open Educational Resources in K–12 Education

Daniel Ribeiro dos Santos¹, Leticia Lopes Leite¹

¹Department of Computer Science – University of Brasília (UnB)

drsantos@gmail.com, l1lleite@unb.br

Abstract. *The reuse, adaptation, and remixing of Open Educational Resources (OER) require mechanisms that preserve attribution, licensing, and derivation history across repositories, schools, and informal sharing channels. This paper proposes a cost-aware blockchain–IPFS architecture for traceable OER publication in K-12 education contexts. In the proposed design, OER packages and provenance-rich manifests are stored in IPFS, while the blockchain records only a compact verification anchor derived from the package identifier. This separation allows authorship, licensing, revision, and remix relationships to remain bundled with the resource without moving descriptive metadata into smart-contract state. A prototype was implemented and evaluated through controlled blockchain and IPFS experiments, comparing the proposed minimal-anchor strategy with a literature-aligned metadata-on-chain baseline. The results show that the proposed architecture keeps blockchain registration cost predictable as the number of co-authors increases, while IPFS supports publication, revision, remixing, retrieval, and provenance reconstruction at the package layer. These findings indicate the use of blockchain as a compact trust layer and IPFS as a provenance-rich storage layer for collaborative OER ecosystems.*

Resumo. *A reutilização, a adaptação e a remixagem de Recursos Educacionais Abertos (REA) exigem mecanismos que preservem a atribuição de autoria, o licenciamento e o histórico de derivação entre repositórios, escolas e canais informais de compartilhamento. Este artigo propõe uma arquitetura blockchain–IPFS orientada a custos para a publicação rastreável de REA em contextos da educação básica. No modelo proposto, os pacotes de REA e os manifestos enriquecidos com informações de proveniência são armazenados no IPFS, enquanto a blockchain registra apenas uma âncora compacta de verificação derivada do identificador do pacote. Essa separação permite que as relações de autoria, licenciamento, revisão e remixagem permaneçam vinculadas ao recurso, sem transferir metadados descritivos para o estado dos contratos inteligentes. Um protótipo foi implementado e avaliado por meio de experimentos controlados com blockchain e IPFS, comparando a estratégia proposta de âncora mínima com uma abordagem de referência, alinhada à literatura, baseada no armazenamento de metadados na blockchain. Os resultados mostram que a arquitetura proposta mantém previsível o custo de registro na blockchain à medida que aumenta o número de coautores, enquanto o IPFS oferece suporte à publicação, revisão, remixagem, recuperação e reconstrução da proveniência na camada de pacotes. Esses achados indicam o uso da blockchain como uma camada compacta de confiança e do IPFS como uma camada de armazenamento rica em informações de proveniência para ecossistemas colaborativos de REA.*

1. Introduction

Open Educational Resources (OER) are intended to support not only access to educational materials, but also their adaptation and circulation under open licenses [Wiley and Hilton 2018]. In practice, however, sustaining these freedoms at scale requires technical mechanisms that preserve attribution across derivation chains, keep revision histories interpretable, and allow resources to remain discoverable over time.

These challenges become more acute in K–12 education, where infrastructure constraints are common. Internet access may be intermittent, bandwidth-limited, or subject to institutional restrictions, and many teachers and students rely on mobile connectivity on personal or shared devices. In such contexts, sharing often migrates from formal repositories to ad hoc channels such as email attachments, messaging applications, and local folders. Although this enables rapid circulation, it also tends to detach resources from their publication context, including their version history, derivation references, and licensing conditions. As a result, different local copies may coexist without a persistent identifier or a verifiable record of provenance.

Blockchain-based approaches have been proposed to strengthen integrity, traceability, and attribution in educational resource ecosystems [Jones 2018]. Yet solutions that place large amounts of data and metadata on-chain face cost and scalability barriers, especially in public networks [Duong and Jun 2021]. This creates a practical tension: the same mechanisms that can improve trust may also raise financial and technical barriers to participation.

To address this tension, this paper presents a hybrid blockchain–IPFS architecture for collaborative OER in K–12 education. The central design choice is to decouple rich OER description from blockchain verification: the complete package remains content-addressed in IPFS, whereas the blockchain stores only a fixed-size identifier derived from the package CID. This separation is intended to preserve verifiability and traceability while keeping on-chain cost predictable and low.

The contributions of this work are threefold: (i) the design and implementation of a blockchain–IPFS prototype for publishing OER packages whose off-chain metadata describe authorship, licensing, versions, and derivation relationships, while publication is anchored by minimal on-chain records; (ii) an experimental blockchain evaluation that compares the proposed design with a literature-aligned metadata-on-chain baseline derived from recurring design patterns observed in recent blockchain-based educational-resource systems; and (iii) a local Kubo/IPFS evaluation of package addition, retrieval, and provenance reconstruction across publication, revision, and remix scenarios.

2. Background and Related Work

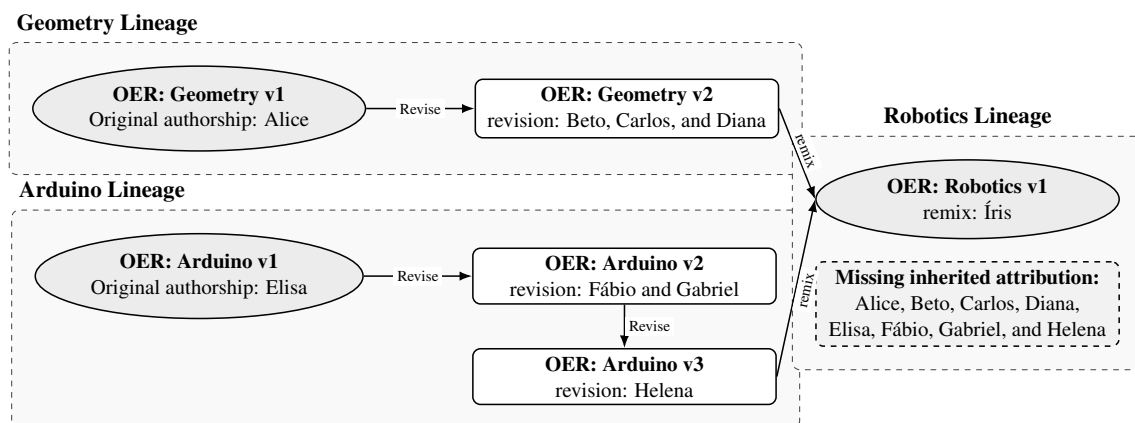
Open Educational Resources emerged from the broader movement toward open access to educational knowledge [UNESCO 2002]. In contemporary Open Education, however, OER are not defined only by free access. Their value also depends on the possibility of retaining, reusing, revising, remixing, and redistributing educational materials under open licenses, commonly referred to as the 5R permissions [Wiley and Hilton 2018, UNESCO 2019]. This perspective shifts the focus from availability alone to the practical conditions that allow resources to be adapted to different curricular, linguistic, and institutional contexts.

For these permissions to be effectively exercised, openness must be supported by operational mechanisms. Open licenses establish the legal basis for reuse, but they do not guarantee that resources will be easy to edit, package, describe, or exchange across platforms. Prior studies therefore emphasize the importance of metadata quality, interoperability, and repository practices for enabling discovery, adaptation, and reuse [Sousa et al. 2023, Menzel 2020, Tavakoli et al. 2021].

A related requirement is provenance. In this work, provenance refers to the record of origin, authorship, institutional context, licensing conditions, and transformations associated with an educational resource. Such information supports attribution, license compliance, transparency, and trust during reuse [Schroeder 2023, Gurung et al. 2023]. When this record is incomplete or fragmented, educators must reconstruct context manually, which increases the effort required for reuse and may contribute to dark reuse, in which adaptations circulate outside repositories and lose their traceable lineage [Beaven 2018, Coughlan et al. 2013].

This issue is especially relevant because OER evolution is not always linear. Revisions generate new variants, while remixing may combine multiple sources into a new resource. As a result, the history of an OER often forms a derivation graph rather than a single version chain. As illustrated in Figure 1, attribution and lineage can be fragmented when these relationships are not consistently preserved across repositories, local environments, and informal sharing channels [Ribeiro dos Santos et al. 2023]. Thus, collaborative OER ecosystems require provenance mechanisms capable of preserving both lineage and attribution across platforms and local copies.

Figure 1. Illustrative OER genealogy and remixing scenario. Source: created by the authors.



2.1. Blockchain for OER

Blockchain has been explored as a tamper-evident layer for registering publication events, resource identifiers, and attribution claims in educational-resource ecosystems [Nakamoto 2008, Jones 2018].

However, the same properties that make public blockchains useful for verification also introduce cost and scalability constraints. Storing large educational files, rich metadata, or detailed histories directly on-chain can increase transaction cost and limit practical

adoption, particularly when public networks are considered \cite{duong_etherlearn_2021}. For this reason, prior systems often adopt hybrid designs. For example, completed teaching materials may be kept off-chain while Ethereum records authorship and contribution information \cite{chou_tmchain_2022}. EtherLearn similarly relies on smart contracts for platform logic, while recognizing that mainnet gas fees may become a barrier to deployment \cite{duong_etherlearn_2021}. These examples suggest that blockchain is better suited to compact verification and accountability records than to serving as the primary storage layer for OER content.

2.2. Blockchain–IPFS Architectures for OER

IPFS is a peer-to-peer, content-addressed file system in which resources are retrieved through content identifiers (CIDs) rather than location-dependent paths [Benet 2014]. Since a CID is derived from the content it identifies, IPFS provides a natural mechanism for checking whether a retrieved file or directory corresponds to the expected content. This property is useful for educational resources that may be copied, mirrored, adapted, or redistributed across different institutional and local environments.

Several blockchain-based educational-resource proposals combine IPFS or other off-chain storage mechanisms with smart contracts. In the OER domain, an Ethereum–IPFS framework has been proposed in which educational hypermedia is stored in IPFS and referenced from the blockchain, reducing dependence on centralized servers [Marjit and Kumar 2020]. More broadly, recent systems show a recurring architectural pattern: files are often kept off-chain, while blockchain transactions or contracts are used to register resource-related information such as ownership, authorship, access rights, timestamps, update events, or copyright evidence.

Table 1 summarizes this pattern in recent blockchain-based educational-resource systems. The table focuses on metadata categories that may increase the on-chain footprint when they are included in function calls, contract logic, or persistent state.

The evidence summarized in Table 1 motivates the baseline used in the experimental evaluation. The baseline does not reproduce a single prior system, nor is it intended as an artificially inefficient comparison. Instead, it abstracts a literature-aligned metadata-on-chain design class, in which provenance-related information is included in the on-chain interface, processed by contract logic, or persisted in contract state. In such designs, the cost of registration may grow with the number of authors, contributors, versions, or recorded events.

The architecture proposed in this paper evaluates a different design choice: keeping rich descriptive and derivation metadata in the IPFS package while registering only a fixed-size identifier on-chain. The comparison is therefore between two plausible architectural strategies for blockchain-supported OER systems: metadata-on-chain registration and minimal on-chain anchoring.

This literature indicates that blockchain can support trust, attribution, and integrity in educational-resource ecosystems, especially when combined with off-chain storage. Nevertheless, existing approaches rarely combine three requirements in a single design: predictable on-chain cost, explicit support for revision and remixing relationships, and suitability for constrained school-network scenarios. This work addresses this gap by

Table 1. Recent literature supporting the metadata-on-chain baseline.

Work	Information causing on-chain growth	A	H	T
[Chen et al. 2025]	Resource registration includes owner/contribution identification, hashing, timestamping, and private-key authentication.	✓	✗	✓
[Silaghi 2025]	The smart contract records resource descriptors and uploader identity, including title, description, IPFS hash, and Ethereum address.	✓	✗	✗
[Rani et al. 2024]	NFT/blockchain metadata binds resources to ownership, authorship-related copyright claims, user details, licensing terms, and block timestamps.	✓	✗	✓
[Li et al. 2024]	Smart contracts support copyright confirmation, ownership tracing, trading, pricing, arbitration, and transaction records.	✓	✗	✗
[Liu et al. 2024]	Resource nodes, contributors, updates, signatures, incentives, and educational activities are handled through smart-contract records.	✓	✓	~
[Liu and Palaoag 2024]	Resource hashes, metadata, uploader/identity information, access rights, transaction behavior, and update hashes are represented on-chain.	✓	✓	~
[Lu 2024]	Shared records, usage-right allocation, copyright management, and policy/key updates are handled through blockchain-supported access control.	✓	~	✗
[Zhao et al. 2024]	Encrypted user/copyright information, multidimensional copyright evidence, infringement verification, and rights-protection records are stored or coordinated through blockchain.	✓	~	~
[Yang et al. 2023]	User roles, user actions, resource modifications, and new blocks for modified resources support traceable resource evolution.	✓	✓	✗
Proposal of this article	The blockchain stores only a fixed-size <code>resourceId</code> , publisher address, and timestamp; authorship, licensing, versioning, and remix provenance remain in the IPFS package.	~	✗	✓

Legend: **A** = authorship, ownership, creator, contributor, uploader, wallet address, or other identity data linked to a resource; **H** = versioning, iteration, update history, action log, copyright maintenance, or resource-history record; **T** = timestamp, creation date, registration date, or block timestamp explicitly associated with publication or registration. ✓ = explicit support; ~ = partial or indirect support; ✗ = not specified.

evaluating whether a minimal anchoring strategy can preserve verifiability without moving provenance-rich descriptions into smart-contract state.

3. Design Requirements

We derive the following requirements from the principles of open education, the 5R permissions, and the operational constraints of K–12 environments.

R1. 5R-oriented workflows. The system must support workflows that operationalize the 5R permissions of open education, allowing resources to be retained, reused, revised, remixed, and redistributed without losing attribution, licensing, or provenance information.

R2. End-to-end provenance. The system must preserve provenance across revisions and remixes, allowing lineage and attribution to be reconstructed beyond the most recent version.

R3. Co-authorship and attribution. The system must support multiple authors and institutional collaborators while keeping attribution associated with both original and derived resources.

R4. Resilience to low connectivity. The system must support creation, access, adaptation, and redistribution of resources in contexts where continuous internet connectivity cannot be assumed.

R5. Cost-aware trust. The system must provide verifiable publication records while keeping participation costs predictable and independent of resource size or provenance complexity.

4. System Architecture

The proposed architecture is organized into two complementary layers: 1. an IPFS layer for storing OER packages and their provenance manifests; 2. a blockchain layer for registering a compact verification record.

In the publication flow, the OER package directory is added to IPFS using fixed IPFS add parameters so that identical content and directory structure produce deterministic CIDs. This process returns the package CID (`packageCid`), that is, the CID of the complete directory, from which the client computes the on-chain anchor:

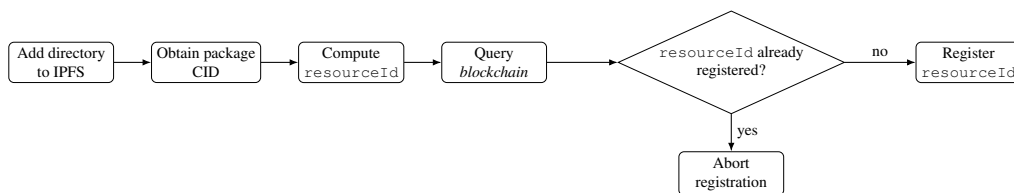
$$\text{resourceId} = \text{keccak256}(\text{bytes}(\text{packageCid})) \quad (1)$$

A fixed-size 32-byte identifier is used on-chain instead of the CID itself because CIDv1 is variable-length. Using the CID directly would therefore make calldata and storage costs variable, whereas `resourceId` preserves a constant on-chain record structure and more predictable cost.

Before sending the transaction, the client checks whether this `resourceId` has already been registered. If it has, publication is aborted; otherwise, the client submits the registration transaction. Thus, full content and provenance remain in IPFS, while the blockchain stores only a compact, verifiable record.

Figures 2 and 3 summarize the registration procedure and show how it supports the requirements defined above.

Figure 2. Verification and registration of `resourceId`

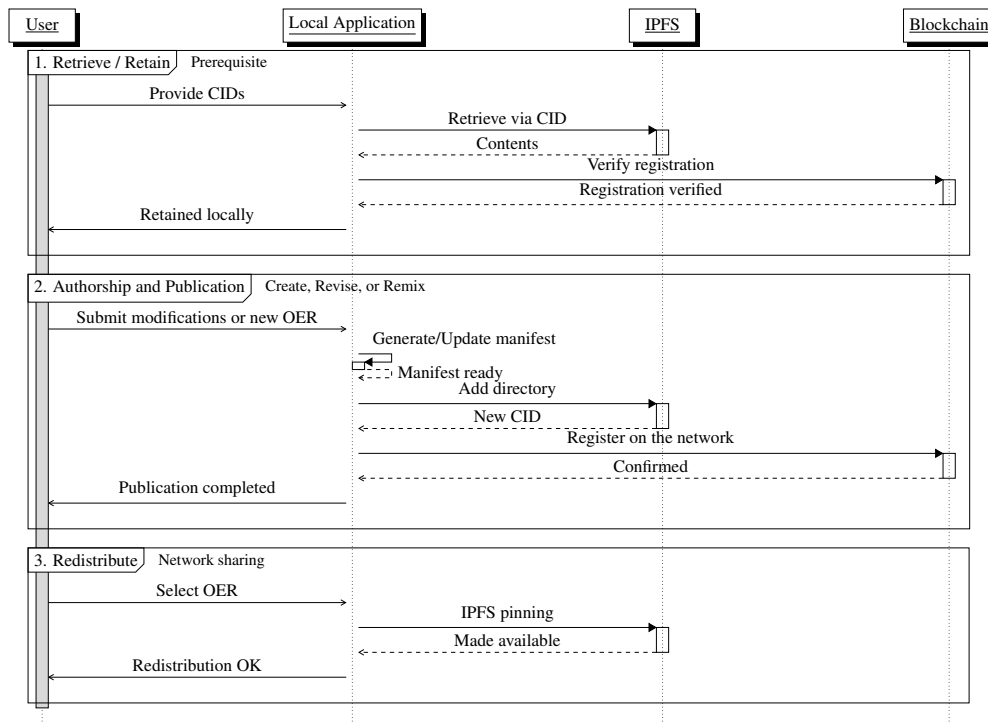


4.1. OER Package Model

Each publication is represented as an OER package stored in IPFS as a directory. This directory contains the educational resource files and a JSON manifest with descriptive and provenance metadata, such as title, curricular component, learning objectives, license, authorship, institution, language, and creation date.

For provenance, the manifest records source package references in the `parents` field and indicates the derivation type in `derivation_type`, distinguishing among NEW, REVISION, and REMIX. The `license_assertion` field preserves the declared

Figure 3. OER Provenance System



license and the associated attribution statement. With this structure, authorship, licensing, and derivation history remain bundled with the resource itself and can be reconstructed directly from the IPFS package.

4.2. Blockchain Trust Anchor

The blockchain trust anchor consists of a mapping from each `resourceId` to a minimal publication record containing the publisher address and block timestamp. This record is sufficient to verify that a given IPFS package was registered, while keeping the on-chain data structure independent of content size, author count, and derivation complexity.

5. Implementation

The prototype was implemented with four components: (i) Solidity smart contracts that implement the minimal on-chain trust anchor and the metadata-on-chain baseline; (ii) an Ethereum Sepolia deployment used to evaluate blockchain registration under a public testnet environment; (iii) Kubo/IPFS nodes for storing and distributing OER directories, with optional local pinning for persistence; and (iv) a Node.js client that interacts with Ethereum and IPFS, performs record verification, publication, and retrieval, and exposes the derivation modes NEW, REVISION, and REMIX.

The blockchain component was therefore evaluated through confirmed Sepolia transactions rather than through a local blockchain simulator. This choice makes the benchmark closer to a public-network deployment scenario, while still avoiding mainnet costs.

5.1. Smart Contract

The smart contract implements the registration procedure for new `resourceId` values. When a publication transaction is submitted, the contract checks whether the identifier is

non-null and has not been previously registered. Valid identifiers are stored and emitted through an event, whereas invalid or duplicate submissions are rejected.

Because registration performs a constant number of checks and stores a fixed-size publication record, its execution cost is $\mathcal{O}(1)$ with respect to OER package size, number of authors, and derivation-graph complexity. This implementation reinforces the architectural role of blockchain as a verification layer rather than as a repository for educational content or provenance metadata.

6. Experimental Evaluation

The evaluation has two complementary parts. First, the blockchain benchmark isolates the cost effect of one architectural decision: whether co-authorship data are processed on-chain during publication or kept in the IPFS package. We compare the proposed minimal-anchor design with the literature-aligned baseline derived in Table 1. Second, the Kubo/IPFS benchmark evaluates the off-chain package layer used by the proposed architecture, focusing on package addition, retrieval, and provenance reconstruction under increasing package size and derivation complexity.

6.1. Research Questions

The evaluation addresses the following research questions:

RQ1: How does gas consumption vary with the number of authors N when comparing the proposed minimal-anchor design with a literature-aligned metadata-on-chain baseline?

RQ2: What transaction fees are observed when the proposed and baseline designs are executed on Ethereum Sepolia?

RQ3: How do IPFS package operations behave as package size and provenance complexity increase in the proposed off-chain storage layer?

6.2. Experimental Environment

The experiments were conducted on Ubuntu 25.10 on a workstation with an Intel i9-13900H CPU, 32 GB RAM, NVMe SSD, and Node.js v24.13.0. The blockchain benchmark was executed on Ethereum Sepolia using a public RPC endpoint. Two contracts were deployed: one implementing the proposed minimal-anchor design and one implementing the conventional metadata-on-chain baseline.

The Sepolia experiment comprised 102 confirmed transactions: two deployment transactions and 100 benchmark transactions. The benchmark transactions covered five co-authorship cardinalities, $N \in \{1, 5, 20, 50, 100\}$, for both implementations, with 10 successful registration transactions per scenario. Each transaction used a distinct `resourceId`, avoiding duplicate-registration reverts and allowing all benchmark transactions to be confirmed on-chain.

For each confirmed transaction, the client recorded the transaction hash, block number, gas used, effective gas price, transaction fee, and client-observed confirmation latency. Execution gas was analyzed separately from intrinsic gas in order to distinguish contract-side execution cost from calldata-related transaction cost.

The IPFS benchmark was conducted with Kubo 0.39.0 through its local HTTP API, using CIDv1 in base32, SHA2-256 hashing, raw leaves, and a 32 KiB chunker. The benchmark used deterministic test data generated from a fixed seed and covered package sizes from 128 KiB to 50 MiB across publication, revision, and remix scenarios. The IPFS benchmark remained local and controlled; it does not measure public IPFS retrieval latency, replication behavior, or long-term availability.

6.3. Blockchain benchmark

Two implementations of the same registration task were compared on Ethereum Sepolia:

- **Proposed design:** the registration operation receives only `resourceId` and stores a fixed-size publication record composed of the publisher address and times-tamp. Authorship and provenance information remain in the IPFS package.
- **Conventional metadata-on-chain baseline:** the registration operation receives and stores co-authorship information during publication. Thus, both calldata size and contract-side execution increase with the number of authors N .

Co-authorship cardinality was used as the workload variable because authorship-related information is a recurring category in blockchain-based educational-resource systems and directly affects metadata-on-chain registration designs. The experiment varied N across five values: 1, 5, 20, 50, and 100 authors. The higher cardinalities are not intended to represent typical classroom authorship, but to stress-test the scaling behavior of metadata-on-chain registration when contributor lists grow.

For each value of N , 10 successful registration transactions were submitted and confirmed for each implementation, resulting in 100 benchmark transactions. The benchmark recorded total transaction gas, execution gas, intrinsic gas, effective gas price, transaction fee, transaction hash, block number, and client-observed confirmation latency. This design evaluates whether moving co-authorship metadata from the on-chain registration path to the IPFS package reduces gas consumption and observed testnet transaction fees.

6.4. IPFS/Kubo benchmark

The IPFS/Kubo benchmark exercised the off-chain package workflow used by the prototype. Each measured publication generated an OER package directory containing a manifest and an educational asset, added that directory to Kubo, recovered the manifest and asset through the resulting CID, reconstructed the provenance graph from parent references, and recorded block-sharing estimates with parent packages.

The benchmark included seven scenarios. Five scenarios varied package content and derivation mode: an original publication (NEW), a metadata-only revision that reuses the previous asset, a content revision that changes one deterministic byte, and remixes with two and three parents. Two additional scenarios stressed provenance metadata by using a small fixed asset and increasing the number of parent references to 10 and 50. These scenarios test whether provenance-rich package operations remain outside the on-chain path while still producing new package CIDs and verifiable derivation relationships.

7. Results

This section reports the results according to the three research questions: gas consumption, transaction latency, and IPFS/Kubo package operations. The findings show the expected

separation between a constant-size on-chain anchor and a provenance-rich off-chain package layer.

7.1. Blockchain Gas Consumption

This subsection reports the Ethereum Sepolia results for the proposed minimal-anchor design and the baseline. The benchmark varied co-authorship cardinality across $N \in \{1, 5, 20, 50, 100\}$, with 10 confirmed transactions per scenario and implementation.

Table 2 shows that the proposed design keeps execution gas constant at 26,959 units and total gas close to 49.6 thousand units for all values of N . This occurs because the on-chain operation receives only a fixed-size `resourceId`. In contrast, the baseline grows with N , since authorship metadata increases `calldata` size and contract-side processing.

Table 2. Gas consumption by scenario (mean $\pm$ std).

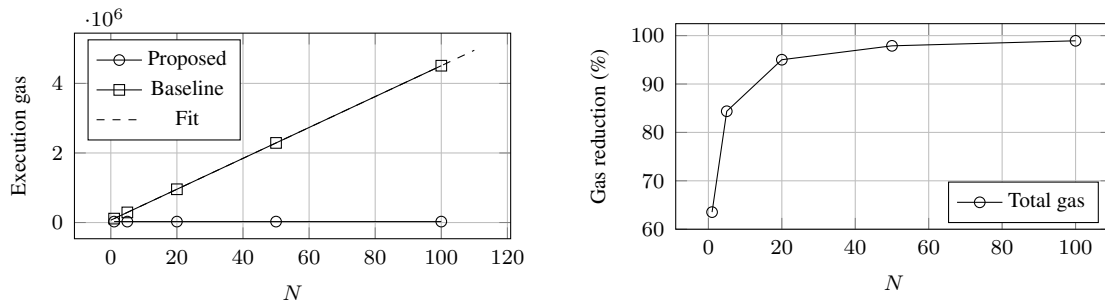
N	Execution gas		Total gas	
	Proposed	Baseline	Proposed	Baseline
1	26,959 $\pm$ 0	112,899 $\pm$ 0	49,573 $\pm$ 5	135,909 $\pm$ 8
5	26,959 $\pm$ 0	290,523 $\pm$ 0	49,571 $\pm$ 6	317,064 $\pm$ 6
20	26,959 $\pm$ 0	956,613 $\pm$ 0	49,573 $\pm$ 5	996,281 $\pm$ 26
50	26,959 $\pm$ 0	2,288,793 $\pm$ 0	49,575 $\pm$ 0	2,354,819 $\pm$ 36
100	26,959 $\pm$ 0	4,509,093 $\pm$ 0	49,573 $\pm$ 5	4,618,985 $\pm$ 70

The baseline execution gas follows a linear model:

$$Gas_{baseline}^{exec}(N) \approx 44,406N + 68,493, \quad R^2 = 1.0000. \quad (2)$$

At $N = 100$, the baseline consumed 4,509,093 execution gas units and 4,618,985 total gas units, while the proposed design remained at 26,959 and 49,573, respectively. Figure 4 summarizes this divergence and the corresponding total-gas reduction, which increases from 63.53% at $N = 1$ to 98.93% at $N = 100$.

Figure 4. Execution gas and total-gas reduction as co-authorship cardinality increases.



Since Sepolia transaction fees are computed from gas used and the effective gas price, fee behavior follows the same structural pattern: the proposed design remains predictable across co-authorship cardinalities, whereas the baseline becomes progressively more expensive as N increases. These values should be interpreted as testnet evidence of transaction-cost behavior, not as mainnet monetary costs.

Overall, the Sepolia results answer RQ1 by showing constant gas consumption for the proposed design and linear growth for the baseline, supporting the architectural choice of keeping authorship and provenance metadata outside the on-chain registration path.

7.2. IPFS/Kubo Package Operations

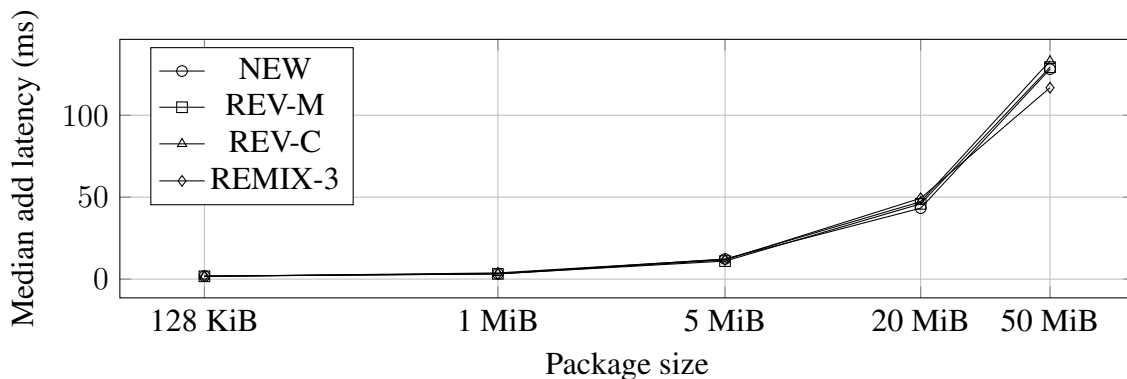
Table 3 summarizes the Kubo/IPFS results for the main publication, revision, and remix scenarios. Figure 5 shows the corresponding median package-addition latency as package size increases. The results answer RQ3 by showing that, in the controlled local environment, package-addition latency scales mainly with content size and remains close across the main derivation modes. For 50 MiB packages, the median addition time ranged from 116.788 ms for the three-parent remix to 133.135 ms for the content revision.

Table 3. Kubo/IPFS package-addition results for selected scenarios (30 runs).

Size	NEW				REV-M				REV-C				REMIX-3			
	p50	p95	Thr.	Sh.	p50	p95	Thr.	Sh.	p50	p95	Thr.	Sh.	p50	p95	Thr.	Sh.
128 KiB	1.826	7.048	67.4	0	1.730	9.548	71.1	2	1.700	5.000	66.9	1	1.826	4.564	70.8	3
1 MiB	3.320	5.053	316.6	0	3.229	11.528	290.0	2	3.790	15.422	277.5	1	3.122	7.059	286.6	3
5 MiB	12.272	21.516	406.2	0	10.993	15.497	452.5	2	12.107	17.225	426.0	1	11.698	24.951	418.2	3
20 MiB	43.285	89.963	443.6	0	45.925	72.258	458.7	4	47.223	67.803	459.8	3	49.347	55.690	428.6	4
50 MiB	128.122	172.891	377.2	0	129.322	155.233	378.7	4	133.135	162.340	376.6	3	116.788	175.318	427.6	6

p50/p95 = package-addition time percentiles (ms); Thr. = throughput (MiB/s); Sh. = shared blocks.
 REV-M = metadata-only revision; REV-C = one-byte content revision; REMIX-3 = remix with three parents.

Figure 5. Median Kubo/IPFS package-addition latency by scenario.



The revision and remix checks also provide evidence for the intended package semantics. The metadata-only revision reused the asset CID from the previous package while producing a new package CID, because the manifest changed. The content revision changed one deterministic byte and therefore produced a new asset and package. Remixes shared blocks with their parent packages, with the three-parent remix sharing six blocks on average in the 50 MiB case.

The full benchmark also included metadata-stress remixes with 10 and 50 parents. Unlike the content-size scenarios reported in Table 3, these cases used a small fixed asset and varied only the number of parent references in the manifest. They should therefore be interpreted as provenance-metadata stress tests rather than as package-size scalability results. Their median package-addition times were 1.376 ms and 1.326 ms, respectively, while the mean repository-size delta increased from approximately 6.1 KiB to 22.0 KiB. This result supports the architectural claim that increasing derivation metadata affects the IPFS package and repository footprint, but not the constant-size on-chain registration path.

7.3. Limitations and Threats to Validity

The blockchain benchmark was conducted on Ethereum Sepolia, which provides a public testnet environment but does not reproduce all economic and operational conditions of Ethereum mainnet or production L2 networks. Sepolia fees are paid in test ETH and should not be interpreted as direct monetary costs. However, gas consumption remains a meaningful proxy for comparing the computational and storage footprint of the two contract designs.

The Kubo/IPFS benchmark complements the blockchain evaluation but remains local and controlled. It measures package addition, local retrieval, provenance reconstruction, and block-sharing behavior through the local Kubo API; it does not estimate public IPFS retrieval latency, replication behavior, pinning availability, or long-term persistence across distributed peers. The generated packages are synthetic and therefore do not replace user studies with teachers working on real curricular materials.

8. Conclusion and Future Work

This paper presented a hybrid blockchain–IPFS architecture for traceable OER publication in K–12 education. The design stores complete resource packages and provenance manifests in IPFS, while the blockchain records only a fixed-size verification anchor.

The Sepolia experiments showed that this strategy keeps execution gas constant and total transaction gas nearly constant as the number of co-authors increases, whereas the metadata-on-chain baseline exhibits linear growth. The local Kubo/IPFS benchmark further showed that publication, revision, remixing, retrieval, and provenance reconstruction can be handled at the package layer without increasing the on-chain record.

In educational settings, this separation may support teachers and educational institutions by keeping attribution, licensing, and derivation references bundled with the resources. Schools can cache and exchange packages through local IPFS nodes during periods of limited external connectivity, while blockchain registration or verification can occur when Internet access becomes available. Persistence may be provided through institutional nodes, shared educational infrastructure, or contracted pinning services. Clients that retrieve packages may also contribute to their availability while they remain online and retain the corresponding blocks. Because the on-chain record remains compact, registration costs become more predictable and easier for institutions to plan as OER collections grow.

Future work includes deployments in school networks, evaluation on production L2 or permissioned networks, governance mechanisms for persistence, identity, corrections, and disputed claims, and user studies with teachers engaged in real revision and remixing workflows.

Declaration on the Use of Artificial Intelligence

The authors declare that Generative Artificial Intelligence tools, specifically ChatGPT, were used as auxiliary support during the preparation of this manuscript. Their use was limited to language revision, grammar checking, and clarity improvement. After using this tool, the authors reviewed and edited the content as needed, and take full responsibility for the accuracy and integrity of the final publication.

All scientific content, methodological decisions, experimental procedures, data analysis, interpretation of results, and conclusions were defined, reviewed, and validated by the authors. The authors take full responsibility for the final version of the manuscript, including its originality, accuracy, integrity, and compliance with the ethical guidelines of SBIE and the Brazilian Computer Society.

References

- Beaven, T. (2018). ‘Dark reuse’: an empirical study of teachers’ OER engagement. *Open Praxis*, 10(4):377.
- Benet, J. (2014). IPFS - Content Addressed, Versioned, P2P File System.
- Chen, W., Tang, Z., and Bai, J. (2025). The application of blockchain technology in multi-modal educational resources. *Journal of Systems Science and Information*, 13(1):137–156.
- Coughlan, T., Pitt, R., and McAndrew, P. (2013). Building open bridges: collaborative remixing and reuse of open educational resources across organisations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 991–1000, Paris France. ACM.
- Duong, T. N. B. and Jun, J. Y. T. (2021). EtherLearn: Decentralizing Learning via Blockchain. In *2021 IEEE International Conference on Engineering, Technology & Education (TALE)*, pages 212–217, Wuhan, Hubei Province, China. IEEE.
- Gurung, I., Adhikari, S., Marouane, A., Pandey, R., Dhakal, S., and Maskey, M. (2023). Exploring Blockchain to Support Open Science Practices. In *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, pages 1205–1208, Pasadena, CA, USA. IEEE.
- Jones, S. (2018). A Solution to OER Publication Resistance: Using Blockchain Technology to Protect Scholar Copyright. *International Journal of Open Educational Resources*, 1(1).
- Li, X., Peng, J., Gao, S., Shi, Z., and Li, C. (2024). Achieving fair and accountable data trading for educational multimedia data based on blockchain. *Wireless Networks*, 30(5):4389–4401.
- Liu, M. and Palaoag, T. (2024). A study on blockchain-based educational resource sharing framework and acceptability. In *IET Conference Proceedings*, volume 2024, pages 80–90.
- Liu, S., Liu, Z., Chen, B., and Pan, X. (2024). Construction and application of on-line learning resource incentive mechanism driven by smart contract. *IEEE Access*, 12:37080–37092.
- Lu, W. (2024). Construction of a secure sharing model for digital educational resources combining blockchain and ABE algorithm in the perspective of smart education. *Informatica*, 48(22).
- Marjit, U. and Kumar, P. (2020). Towards a Decentralized and Distributed Framework for Open Educational Resources based on IPFS and Blockchain. In *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*, pages 1–6, Gunupur, India. IEEE.

- Menzel, M. (2020). LOM for Higher Education OER Repositories. *LOM for Higher Education OER Repositories Ein Metadatenprofil für Open Educational Resources im Hochschulbereich zur Förderung der Interoperabilität von OER-Länderrepositorien*, pages 1–10.
- Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.
- Rani, P., Sachan, R. K., and Kukreja, S. (2024). Educopyright-chain: an educational resources copyright protection system utilizing permissionless blockchain and non-fungible tokens. *Peer-to-Peer Networking and Applications*, 17(6):3583–3602.
- Ribeiro dos Santos, R., Prado Santos, M. T., and Ciferri, R. R. (2023). ProvOER model: A provenance model for Open Educational Resources. *Heliyon*, 9(2):e13311.
- Schroeder, N. (2023). Version Management in a Distributed Infrastructure for Open Educational Resources. In Otto, D., Scharnberg, G., Kerres, M., and Zawacki-Richter, O., editors, *Distributed Learning Ecosystems*, pages 241–261. Springer Fachmedien Wiesbaden, Wiesbaden.
- Silaghi, D. L. (2025). Blockchain-based solution for protecting teachers’ copyrights on educational resources. In *2025 18th International Conference on Engineering of Modern Electric Systems (EMES)*, pages 1–6.
- Sousa, L., Pedro, L., and Santos, C. (2023). A Systematic Review of Systematic Reviews on Open Educational Resources: An Analysis of the Legal and Technical Openness. *The International Review of Research in Open and Distributed Learning*, 24(3):18–33.
- Tavakoli, M., Elias, M., Kismihók, G., and Auer, S. (2021). Metadata Analysis of Open Educational Resources. In *LAK21: 11th International Learning Analytics and Knowledge Conference*, pages 626–631, Irvine CA USA. ACM.
- UNESCO (2002). Forum on the Impact of Open Courseware for Higher Education in Developing Countries, UNESCO, Paris, 1-3 July 2002: final report - UNESCO Digital Library. Technical Report CI.2002/CONF.803/CLD.1, UNESCO.
- UNESCO (2019). Recommendation on open educational resources (oer). Technical report, UNESCO, Paris.
- Wiley, D. and Hilton, III, J. L. (2018). Defining OER-Enabled Pedagogy. *The International Review of Research in Open and Distributed Learning*, 19(4).
- Yang, L., Fu, Y., and Wang, T. (2023). International chinese learning resource management model based on blockchain technology. In *2023 IEEE 3rd International Conference on Electronic Communications, Internet of Things and Big Data (ICEIB)*, pages 278–281.
- Zhao, G., He, H., Di, B., and Guo, Q. (2024). Bc-dercp: Blockchain-based copyright protection mechanism for digital educational resources. *Education and Information Technologies*, 29(15):19679–19709.