Beyond Blockchain:

Policy Brief on Next Generation Web Technologies and their impact on OER

Published in 2024 by the United Nations Educational, Scientific and Cultural Organization,
7, place de Fontenoy, 75352 Paris 07 SP, France

Cover photo: VideoFlow/shutterstock.com

CI-2024/WS/2

Beyond Blockchain:

Policy Brief on Next Generation Web Technologies and their impact on OER

Table of contents


shutterstock.com/gorodenkoff

Introduction

This policy brief follows on the study 'Blockchain for OER: Distributed Ledgers, Self-sovereign Identities and Value Paradigms for Open Education,' produced in 2021 under contract for UNESCO. This work is intended to update that work by analysing the implications of next-generation web technologies especially those linked to blockchain (such as self-sovereign identity, distributed ledgers, verifiable credentials and cryptographic verification) as well as artificial intelligence and OER. In particular, it focuses on:

- The implications of cryptography for attribution and copyright tracking, collaborative content creation, credentialing, certifications, interoperability and interconnected ecosystems.
- An examination of potential benefits but also challenges to address considerations for the successful implementation of distributed trust technologies in the OER landscape.
- The challenges and opportunities of ledgers for the use of artificial intelligence (AI) to support inclusive and equitable access to knowledge sharing and knowledge building in an OER framework.

Background

Blockchain for OER

The 2021 study identified a pressing need within the global Open Educational Resources (OER) community for the establishment of a comprehensive global network dedicated to tracking the publication, usage, and reuse of OER materials. This envisioned network would have several critical functions, including the ability to:

- Identify newly published OER.
- Pinpoint the authors of these resources.
- Clarify the license conditions under which these OER are published.
- Document each instance where an OER is cited or reused in new works.

Despite strong demand for such a service from community members, challenges related to governance and the absence of a viable business model have hindered the development of such a network.

The study identified distributed ledger technologies such as blockchain as a potential solution to address these critical functions.

Blockchain technology introduces the possibility for a decentralised structure where data isn't stored in a single location and is instead distributed across a network. This decentralisation reduces the risk of data monopolies, making it thus easier to distribute and access OER more widely.

Since each block of data is linked to the previous one and distributed across multiple nodes, unauthorised alteration of data becomes computationally impractical, significantly enhancing data security. Since each transaction on the ledger is signed, OER can be timestamped with a time of registration and linked to the author of the transaction.

Blockchain technologies do not require centralised governance by a single entity, allowing instead for a collaborative effort among a diverse group of stakeholders, potentially including those who are otherwise untrusted or would have otherwise difficulty accessing or distributing OER under a centralised or a single entity governance structure.

By enabling direct peer-to-peer transactions without the need for intermediaries, it also gives content creators greater control over their intellectual property and opens new opportunities to receive academic recognition or other forms of compensation for their work.

Since no single stakeholder has the scope, coverage or funding to create the applications needed by the OER community, especially in terms of databases for tracking reuse, distributed ledgers[1]

1 Distributed ledgers are databases that are collectively maintained and updated across various locations, organizations, or regions, and are open to access by numerous individuals. Each network node participant has access to the records distributed throughout the network and possesses an exact duplicate of them. Modifications or new entries to the ledger are quickly replicated and shared with all participants, typically within seconds or minutes. Blockchains are a specific example of distributed ledgers.

were suggested as a way of sharing responsibilities and infrastructure. A globally distributed infrastructure would offer the durability and resilience required to ensure the system lasts longer than any individual project or organisation.

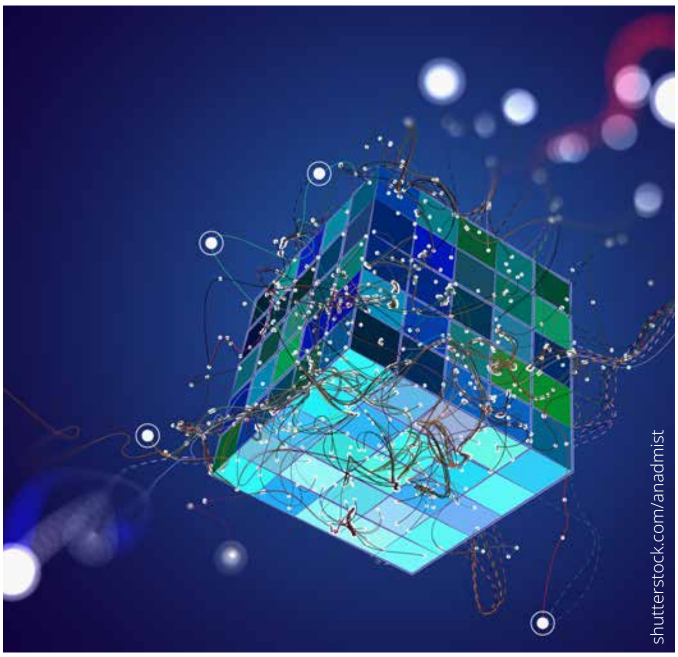
Unmet Promises: The Downfall of Blockchain

The initial excitement surrounding blockchain predicted a sweeping disruption across various industries, including the education sector. However, the real-world deployment of this technology has often fallen short of these grand visions. The excitement outstripped the technology's actual maturity, resulting in inflated expectations about its capabilities and societal impact.

In fact, despite hundreds of blockchain startups having been launched over the past decade, there are still no mainstream products in areas linked to OER such as authentication of documents, intellectual property management or reuse tracking. Many products that started as distributed ledger products have evolved away from blockchains as their underlying technology. According to Crunchbase (a data provider on IT companies and startups), there are no such blockchain companies in its top 1000 companies, except for those linked to decentralised finance or venture capital.

The difference between the ambitious goals advocated for by its supporters and the reality of its current usage requires a careful re-evaluation of both the possibilities and drawbacks of blockchain technology for OER.

One of the core technical challenges encountered by blockchain is scalability. Prominent blockchain networks in the financial industry, such as Bitcoin and Ethereum, despite their popularity and financial backing, struggle to process the significant volumes of transactions efficiently. The inability to solve this issue has left blockchain-based networks significantly slower and more expensive to operate than traditional databases. Given that the education industry tends to receive comparatively lower levels of funding, the cost aspect cannot be ignored when thinking of blockchain technology as an option for OER.



The environmental footprint of blockchain networks, particularly those relying on Proof of Work (PoW) consensus mechanisms, has remained a significant point of contention, with technical fixes proving insufficient to address their substantial energy consumption.

From a policy standpoint, blockchains have quickly gone from a technology to be promoted to one to be heavily regulated to protect citizens. The past decade has seen the blockchain and particularly the cryptocurrency domains witness a surge in fraudulent activities and scams. The prevalence of fraud was particularly pronounced during the Initial Coin Offering (ICO) craze which peaked in 2019, where the lack of regulatory oversight enabled a proliferation of scams, leaving investors with worthless tokens and significant financial losses and the blockchain technology with a bad image. Another round of bad headlines was generated in 2022-2023, with the largest blockchain company[2] in the world going bankrupt due to fraud, and several blockchain CEOs being jailed. Increasingly, this bad reputation has come to overshadow technical discussions. In particular, industries with a strong public or social mission, do not want to associate themselves with technologies with a reputation for financial impropriety.

Further compounding these issues is the notably high failure rate of blockchain startups. Various factors, including the absence of viable business models, the struggle to achieve product-market fit, or the sheer pace of technological evolution, have contributed to an environment where, according to Gartner[3] and Coin Telegraph[4] up to 95% of blockchain projects have faltered or face obsolescence.

Despite its noble intentions and transformative vision, blockchain as a technology has exhibited significant limitations. Its struggles have painted a complex and often discouraging picture of its practical utility and prospects.

In terms of applicability to OER, while the technology still may show promise, there is no longer sufficient political momentum behind its development to warrant any hope of it being able to meaningfully bring benefits to the OER movement in the short-to-medium term.

Impact of Next Generation Web Technologies on OER

The failure of blockchain technology to provide sufficient political and technical momentum to deliver the functionalities required by the OER community, and the education sector more broadly, prompts a look into emerging trends and technologies that could eventually solve these issues.

The emergence of Web 3.0

Despite falling short of expectations, much of the hype around distributed ledger technologies like blockchain stemmed from the recognition that they aimed to address immediate and significant issues with the structure of today's internet, including for the education sector and in particular for OER.

General concerns over the lack of trusted identities, data monopolies, centralisation and worries about privacy and data security are all subjects that are closely linked to the values of the OER community and why blockchain technology was deemed an appropriate response to address those challenges.

Looking forward, a solution that researchers and engineers are now exploring is to integrate alternative technologies and frameworks that can fulfil the need for data trustworthiness and verifiability without relying solely on blockchain. This Web 3.0 movement is a new paradigm that seeks to realise the ideals initially aspired to by blockchain but with a broader and more robust technological foundation. Web 3.0 aims to create a more decentralised, user-centric, and intelligent internet as a whole. It addresses the fundamental flaws of the previous web iterations by fostering an environment where users regain control over their data, identities, and digital interactions.

2 https://www.reuters.com/technology/crypto-companies-crash-into-bankruptcy-2022-12-01/.
3 American technological research and consulting firm.
4 Website news about crypto industry.

There are significant arguments about the difference between 'web3' and 'web 3.0'. While web 3.0 was coined by Tim Berners-Lee as a vision for a semantic, interconnected web, web3 coined by Gavin Wood, the co-founder of Ethereum, is more focused on de-centralisation, especially using blockchain based protocols.

Another perspective suggests that the future iteration of the web will integrate elements of both these concepts. Web 3.0 and web3 together promise to represent a major change towards an improved internet, which would also be relevant for use cases linked to OER. For simplicity we refer to these concepts taken together as Web 3.0. While there is no single definition of Web 3.0, nor a definite list of technologies which contribute to Web 3.0, most definitions will include some reference to decentralisation, enhancing data with semantic meaning, assuring user identity and providing cryptographic assurance of transactions.

In the following, we will discuss some of these future technologies that could be part of Web 3.0 but also operate as standalone technologies. We will also highlight how they might connect to challenges facing OER.

Semantic Web and Artificial Intelligence

The Semantic Web represents a fundamental shift in how data is structured and utilised on the internet, envisioning an ecosystem where data is not just abundant but semantically rich and machine-readable. This paradigm enhances the web's usability and intelligence by enabling machines to understand and process the meaning of data, much like humans do. This would tackle one of the major issues with OER, where data such as authorship and licensing conditions are inherently detached from its content and referencing to previous versions cannot easily be tracked without a central repository.

Artificial Intelligence (AI) and machine learning technologies significantly enhance the web's capability to not only comprehend the meaning of data but also to process and interconnect it in novel and innovative ways. This combination creates a new era in data interaction, where AI systems can analyse and interpret data similar to humans, but much faster and on a larger scale. These advanced technologies can detect patterns, predict outcomes and provide reliable insights with good and continually improving accuracy.

Identity and Ownership

A core tenet of Web 3.0 is the empowerment of users and content creators with full control over their identity and digital assets. Decentralised identifiers (DIDs) and verifiable credentials are emerging as powerful standards in identity management, delivering a decentralised, secure, and user-centric approach. Linking content such as OER to its author(s) and reliably identifying users, while at the same time strengthening user privacy and control will enhance the integrity and authenticity of digital interactions and assets.

Blockchain-linked technologies help in creating a verifiable audit trail for every digital asset, ensuring that the origin, ownership, and history of transactions are indelibly recorded and easily traceable. This level of traceability is useful for protecting intellectual property rights, enabling creators to assert ownership and receive rightful recognition and compensation for their work.

Cryptography and Trust Infrastructures

Blockchain technology laid the foundational stone for a new trust framework by enabling cryptographic assurance of all database entries. Every transaction on a blockchain is secured through advanced cryptographic algorithms, ensuring integrity, authenticity, and non-repudiation. This level of security means that any entry in the blockchain ledger is tamper-evident and traceable, providing a high degree of trust in the data. However, blockchain's capabilities predominantly revolve around anonymised identities and transactions - one of the reasons it became popular as a platform for fraud.

Web 3.0 upholds the principles of cryptographic assurance of transactions, and a high-level decentralisation, but also maintains the idea that transactions must be linked to verifiable identity attributes of users. This concept is visible in technologies from Know Your Customer (KYC) procedures employed by banks to accreditation registries which establish trust in the issuers of digital diplomas.

A Critical Look at Decentralisation

While decentralisation is presented as a key aim for web 3.0 technologies, the past decade has actually seen an acceleration in the centralisation. Blockchain and web 3.0 have had a negligible impact on this trend so far. Commercial cloud providers have dropped prices and increased convenience to the point that they have become the de-facto standard way of deploying web infrastructures, since decentralised infrastructures cannot compete on price. This has led to more and more applications moving to the cloud, and even services such as chat which used to be decentralised, have moved nearly entirely to centralised models.

The AI boom has only increased this trend, with AI models requiring large data centres and massive datasets to enable their training. At the current point in time, the market has decided to favour centralised data processing infrastructures, with over 91% of businesses using cloud solutions according to Rightscale.

Many of the goals for which decentralisation was promoted are instead achieved through different means. Rather than ensuring user control over data through decentralisation, this is achieved through better application design.

End-to-end encryption allows users to communicate securely without the central nodes being able to read the data being transferred. Edge computing promotes the idea that more data should be processed on a user's device, rather than in the cloud, e.g. by doing voice-to-text transcription on a device, rather than by sending voice clips to the cloud.

Experimental technologies such as solid pods allow users to decide which cloud they want to store their application data in, and seamlessly transfer it between clouds (or even their own devices). Secure data wallets which may be stored on-device with encrypted backups in clouds allow users to prove their identity, store their credentials and present them to access services across the internet.

As a result, while many open web advocates still campaign for an open web, the focus of many initiatives now has shifted to ensuring data privacy, interoperability and security atop a cloud-based infrastructure. In short, building an open web on top of a closed infrastructure.

New challenges

As digital assets are increasingly used and ingested by machine-learning models, the importance of maintaining a transparent and accurate record of asset usage becomes even more pronounced. AI models have been trained on massive amounts of publicly available data - GPT3.5 was trained using 570GB of datasets and over 300 billion words of data. Significant questions remain as to whether intellectual property rights were respected in such data collection, and whether this constitutes fair use. Either way, such large-scale implementations only emphasise the need for an ability to track how data is used by AI and Machine Learning (ML) models, ensuring that the use of data adheres to the stipulated terms and conditions set by the creators or owners.

A healthy OER community requires that the rights of creators be protected to foster a fair and equitable ecosystem where contributions are duly acknowledged and rewarded.

Within the context of generative Ais, facts are not so much 'retrieved' but are 'generated' based on patterns of data. It is fair to say that the output of a large language model (LLM) is 'based on' its input data, but it typically cannot be said that it is 'citing' its input data. This indicates the need to re-thinking what the concept of attribution means in the age of AI. This fundamental distinction then has implications as to whether AI models (or their outputs) are bound by the copy-left licences much OER use.

A Web of OER - an updated vision

Independent of whether a decentralised vision of Web 3.0 or a more centralised approach to the web will prevail, the latest developments in technology do not change the overall needs of the OER community. The relevance of blockchain in OER remains to:

- Identify the OER that are published.
- Identify their authors.
- Identify the licence conditions attached to the publication.
- Record each citation or reuse of an OER in a new work.

Using current and emerging web technologies, we can design a 'Web of OER' where traceability of OER is ensured and where discovering relevant open educational resources is easy. This has the potential to lead to a massive increase in the publication, use and re-use of OER, and an overall re-energisation of the OER movement as a whole.

Identification of OER

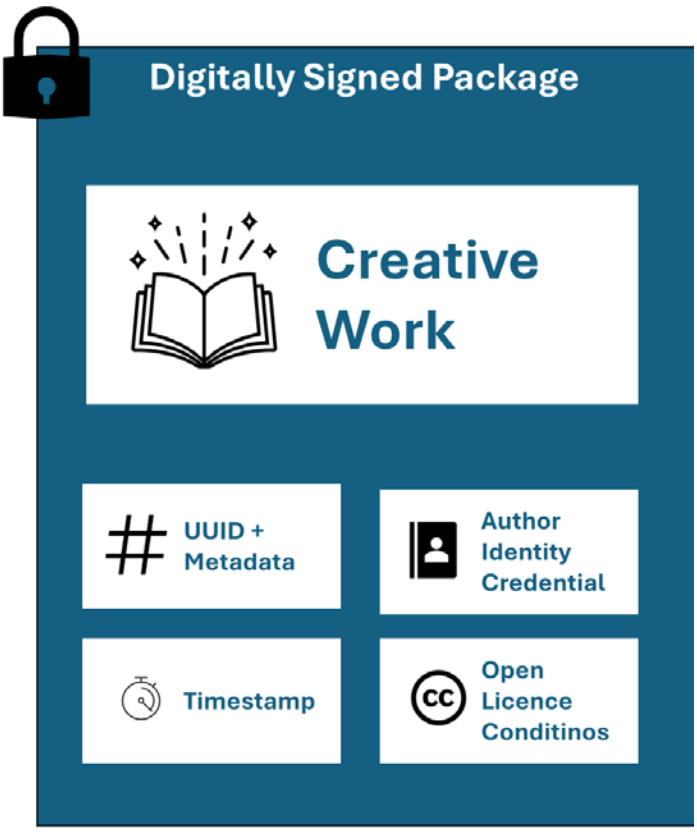
OER can be uniquely identified via a freely available universal identifier such as a UUID (Universally Unique Identifier), together with associated meta-data. UUIDs follow a decentralised principle, since they can be generated by any machine without a central server, while still maintaining uniqueness. The meta-data can follow any number of academic standards of publication such as Dublin-Core and associated standards.

Registering rights of authors

In the Web of OER, every creative work should be unambiguously associated with its authors, with authors able to attach licence conditions to their work, in a tamper-proof and verifiable fashion. Technologically, this is made possible by:

- **Secure digital identities** – numerous technologies and systems now link web handles to real persons. National identity systems such as the EU Digital Identity Wallet in Europe, or Aadhar in India are examples of government-issued IDs. Academic identification systems such as ORC.ID provides identification for researchers. Many web services now require onboarding processes that include verification of identity via the supply of a credit card, phone number or other personally identifiable handle. All of these services provide users with a way of proving who they are and protection against identity theft.
- **Digital signatures** – whether using Public Key infrastructures, blockchain-based signing or even experimental quantum cryptography techniques, digital signatures allow for the cryptographic binding of a creative work with a specific identity, licence and timestamp.

A digitally signed open educational resource



While these technologies and principles have existed for decades, they were extremely hard to use, and often very expensive (especially to verify identity). The last few years have seen the commodification of identity and signature technologies, partly thanks to the decentralised web movement. It has become commonplace for any person to be able to access signature technologies and digital identity documents for little to no cost - with many countries providing these free of charge to all citizens.

Discovering and Re-Using OER

Up until now, policies for improving the discoverability and re-use of OER have focused on improving the metadata which is used to describe resources. Accurate metadata formed the basis of search indexes that allowed documents to be found via keyword searches. More powerful search techniques would allow searching for keywords across the text of documents. Under this paradigm, the purpose of search was to locate relevant resources, then it was the role of a user to parse through these resources to locate useful information.

Semantic (AI-powered) search upends this paradigm. While metadata is still essential to highlight important parts of an OER and help classify it, large language models (LLMs) are able to ingest the entire contents of an OER - be it text, images, audio or video and extract meaning from the text that has been entered. This means that instead of searching for resources, users can query an AI interface such as a chatbot directly for information, and have that information produced. It is now possible e.g. to ask for a short video to be generated explaining a topic based on multiple sources, fully automatically.

These new capabilities break down the barriers between 'search', 'use' and 're-use' given that with the power of new models, the act of requesting information can lead to the creation of new work to understand it. Within this context, it also becomes less relevant to refer to individual resources, and rather to think of all discoverable OER as a single corpus of material.

Tracking Re-Use of OER

An essential incentive for educators to produce OER is the ability to derive value from publishing high-quality OER. While generally OER use and re-use are not monetised, educators and learners who can prove they create popular OER may gain various benefits, including enhanced professional recognition, career advancement prospects, opportunities to be invited to author other OER etc.

For this data to be available, two related processes must be in place:

- Creators of OER need to reference the creative works they are re-using for their new work.
- Services need to exist that aggregate and classify those references, in a similar way to how citations are tracked for academic journals.

The current method of accomplishing this has significant limitations. Manually adding citations to their work is time-consuming and labour-intensive. Moreover, such citations do not refer to how much of a work has been re-used, the size of the intended audience, or to the quality of the new work. Due to a lack of standards for citations in OER, there are no mainstream services that aggregate this information on re-use, making it hard for a creator to track how their work has been adapted or re-used.

Through a combination of cryptographic signing original works and AI, several new possibilities arise. Assuming that the entire corpus of OER is ingested into a single AI model, it is possible to:

- Create AI assistants that would aid in the generation of citations.
- Automatically calculate the degree of re-use in a document based on text analysis.
- Analyse the entire corpus to track trends in citation, use and re-use of data to generate analytics on what is most used.
- Link the data to the authors of the original works, thanks to cryptographic signing, to track the co-creation flow from the initial creator through its various iterations.

Limitations

The described technologies present opportunities to improve attribution and discoverability of OER, but their effectiveness relies on being accompanied by appropriate organisational processes. These technologies by themselves are not capable of ensuring compliance with licence conditions. Digital Rights Management systems are generally ineffective and generally require restrictions that contradict the very principles of the OER community. While AI has the potential to detect the use of licenced content even where it has been adapted, such systems by nature cannot be perfect. Given the multitude of sources describing the same content, it becomes challenging to ascertain which source was utilised to create a specific piece of content.

AI technologies also vastly increase the amount of 'original' content that can be generated by any individual. Many open-access repositories have already been flooded by AI-generated submissions, forcing them to change their upload policies, or in some cases, suspend them entirely. While digital signing of content would allow repositories to identify and blacklist chronic abusers, repositories are only just starting to understand how to deal with this new type of threat.

Furthermore, these technologies fail to address the spread of misinformation. AI and semantic technologies make it trivial to spread misinformation, and to create a network of linked documents citing each other to create virality and to give an impression of accepted facts where there are none. The openness of OER provides a fertile ground for these kinds of threats, and up to now, no compelling technological solutions have been proposed to combat such threats.

Conclusions & Recommendations

R1: 'Crypto' Technologies can play a significant role in promoting OER

While distributed ledgers may not be able to deliver on the promise, they once held to promote OER, linked technologies such as cryptographic signing, semantic-based interoperability and machine learning all hold significant promise in being able to promote the twin aims of improving attribution and discoverability of OER. While none of these technologies is currently commonly used in the OER space there are examples of mature implementations of them in multiple sectors, including examples of them being used in curriculum planning, educational chatbots and qualifications comparison.

R2: Signing Creative Works is Essential to OER Traceability

Using cryptographic signatures to embed authorship and licensing information into OER is essential for ensuring traceability of OER. The technologies required for this purpose are commonly available and mature.

Future updates of OER strategies should integrate this requirement into policies for the production of OER, quality criteria for OER, as well as be considered an essential criterion for inclusion in OER directories. Additionally, education on OER needs to be revised to integrate this principle. Wherever possible, policies should emphasise the connection of signatures to real-world identity of authors - both to create incentives for publication and to counter misinformation efforts.

R3: Global Interoperability standards for traceable OERs are needed

Linked to the recommendations in R2, global standards for OER should move beyond mere bibliographic standards to more technical standards that describe formats for embedding metadata into OER, identifier generation standards, author-identity credentials, time-stamping mechanisms and methods for signing OER packages. The creation of such standards would significantly enhance the discoverability and traceability of OER and lower the barrier to developing software applications that use OER.

R4: Research is required into next-generation citation systems

In light of the advancement of AI technologies, particularly large language models (LLMs), there exists considerable potential in terms of tracking the use and re-use of OER. AI offers the potential of being able to trace the influence of ideas across networks of resources and people, rather than merely track the use of the objects that contain those ideas.

However, up to now, little progress has been made in this domain, except for some experiments using generalised LLMs. Some work has also been done to ensure that AIs such as chatbots cite their sources, but large-scale analysis of connections between works is in its infancy.

Significant research needs to be done to elaborate a concept of what tracing ideas rather than resources would mean, how to utilise AI to do it, and how to present this tracking in ways that are meaningful for the community.

R5: Open-Licences need to be updated to take into account AI and LLMs

There is no technological solution to moderating the use of OERs as training data for LLMs. Open licences need to be updated or clarified to indicate whether such usage counts as 'use' under these licences, and how generative AIs should attribute the use of such training data. Furthermore, copy-left provisions need to further clarify whether they apply to (a) AI models, and (b) the output of such models, when they are trained using such data in those licences.

R6: Implementing Technologies, the right way requires mobilising the entire community

There is no single stakeholder who can take the lead to make such a service become a reality. It requires coordination and action from the entire community. To successfully bring these ideas to production would require the cooperation of:

- R&D Centres to develop the technologies and ideas required for better tracing and tracking, especially using Ais.

- OER Repositories to develop and implement policies on prioritising digitally signed works, and on how their work may be re-used in training of AI models.
- Managers to make references to new systems of identifiers and tracking in policy and training materials.
- Practitioners to make use of the services.
- Governments and private foundations to provide funding for research, development and maintenance of technology systems.

While mobilisation on such a scale poses a significant challenge to the adoption of any system, the confluence of new emerging capabilities, reduced costs and tested approaches provides a unique opportunity for OER to make a qualitative leap forward in adoption.

About the author

Anthony F. Camilleri is the founder and senior partner at the Knowledge Innovation Centre, with 15 years of experience at the intersections of technology, education, and quality assurance. Over the past five years, he has specialized in micro- and digital-credentialing, working on projects to support efficient and flexible learning pathways for students and workers. He has consulted with the European Commission on initiatives like European Digital Credentials for Learning and the European Blockchain Services Infrastructure. Anthony is dedicated to enhancing access to education and employment, believing in the power of education to realize human potential.

OER

For further information please contact:
OER Dynamic Coalition Communication and Information Sector
UNESCO Headquarters, Paris

UNESCO OER page at:
https://www.unesco.org/en/open-educational-resources

Stay in touch:

oerrecommendation@unesco.org

Open Educational Resources | UNESCO
Home | UNESCO OER Dynamic Coalition Portal