Content Authenticity and Provenance in the Age of Artificial Intelligence: A Call-to-Action for the LAMs Community

A Product of the C2PA for G+LAM Community of Practice

February 2026

AUTHORS

Kate Murray

Joshua Sternfeld

CONTRIBUTORS

David Cirella

Ann Hanlon

Nick Krabbenhoeft

Eric Lopatin

In the spirit of this report, we invite community feedback: c2pa@loc.gov

Executive Summary

The libraries, archives, and museums (LAMs) community is entering into a decisive moment as artificial intelligence (AI), especially generative AI, is fundamentally reshaping how collections are created, modified, described, and reused. Fast-changing AI technologies can introduce extraordinary opportunities, but also serious risks. AI may automate some workflows, but it can also challenge long-standing methods for documenting provenance, verifying authenticity, and sustaining public trust across the digital preservation lifecycle. In the face of rapid changes in technology and user expectations, LAMs must address how to ensure that digital collections content remains authentic, transparent, and verifiable from creation through access in order to meet the central community's mission of public trust.

While content authenticity and provenance (CAP) have long been archival principles, existing processes are increasingly impacted, or have the potential to be impacted, by AI-mediated workflows. There are growing expectations from researchers, donors, the public, and heritage practitioners to document these impacts comprehensively and consistently by extending traditional content authenticity and provenance data.

The impact of AI on CAP imposes numerous risks that demand immediate collective attention. Responding to the myriad ways in which AI might threaten trustworthiness will undoubtedly place strains on institutional capacity, including, but not limited to, upgrades to hardware and content management systems, staff training, the emergence of a new class of AI-native positions, and dependencies on industry technologies. Even with the best intentions of maintaining transparency with CAP data, AI technologies introduce novel ethical, legal, and privacy threats, especially for traditionally underrepresented communities. At the same time, AI is transforming in real-time the creation, organization, and analysis of data at a pace that defies the LAMs community's traditionally deliberative response to change.

This report serves as a call-to-action, especially for preservation administrators and practitioners across the LAMs sector, in response to the common question: What should my organization do? Rather than advocating for or against AI use in LAMs, it calls for professionals to educate themselves about the challenges that lie ahead and to develop pragmatic responses based upon institutional capacity and unique circumstances.

The collective call-to-action is organized around four mutually reinforcing pillars.

> **Research and Development:** The LAMs community must invest in sustained R&D that extends established content authenticity and provenance principles to the new AI realities, while ensuring that humans remain meaningfully in the loop across workflows and institutional contexts of all sizes.
>
> **Partnerships and Collaboration:** LAMs must deepen cross-institutional and cross-sector collaborations to avoid duplication, accelerate learning, and co-develop shared frameworks, tools, and standards that reflect diverse digital preservation needs.

Advocacy with Industry, Vendors and User Communities: LAMs should actively shape standards, specifications, practice models, and technologies by asserting their requirements for open, vendor-agnostic, and trustworthy CAP data, leveraging their unique authority as long-standing stewards of public trust.

Open Distribution of Results and Lessons Learned: The rapid pace of AI innovation demands more transparent and collaborative modes of sharing experimental approaches and outcomes to complement traditional scholarly channels.

Taken together, these four pillars offer a pragmatic roadmap to navigating the technological and human-centered disruptions that AI brings while continuing to support foundational preservation principles. There is no single standard tool, standard, or institution that can resolve AI's impact on CAP. Instead, safeguarding robust and achievable CAP options in the age of AI will require creative strategies centered upon human judgement and collaboration. Through collective action the LAMs community can continue to serve as a trusted anchor even as the nature of digital content itself is transformed.

Introduction

The rise of artificial intelligence, especially generative AI, is poised to cause profound and transformative shifts for libraries, archives, and museums (LAMs) in maintaining their shared mission as trustworthy and transparent stewards of information and knowledge.

With the stroke of a keyboard, the click of a mouse, or execution of a script, AI can create new content or change existing content at any level – from a fraction of a pixel in an image to an entirely AI-generated video. Changes can alter the integrity or apparent truth of the content in ways known and obvious, but also hidden and obscure. This fast-changing technology has advantages in terms of expanded productivity, but it brings its own challenges in how to document and verify the provenance and authenticity of a collection's content over its lifespan.

For example, if you are an archivist receiving a set of photographs from a photojournalist documenting a newsworthy event, how can you verify that the images were not altered by AI in such a way as to distort what occurred? If you're a researcher studying a digital capture of an ancient artifact, how can you confirm that the capture device and settings are accurately documented? If you're a special collections librarian receiving a set of chatbot transcripts illustrating an author's creative process, how should you document its provenance? If you are indexing and transcribing a recorded interview, what measures should be taken to guarantee that the voices are authentic and words or whole passages not synthetically inserted or deleted? If you're a documentary filmmaker, what steps should you take to distinguish historical from AI-generated content?

These scenarios, and countless others, convey what is at stake for institutions entrusted with preserving our collective heritage. Every administrator faces a version of the same question: How does my institution ensure the integrity of its collection when there are so many points along the preservation lifecycle in which AI technologies could disrupt authentication?

Administrators and heritage practitioners from all institutional types and sizes comprise the primary audience for this high-level report, which outlines the stakes for authentication in an AI-saturated environment. But first, a brief word about what this report is not: We do not survey the growing landscape of LAMs' experimentation with AI tools. It is certainly the case that AI offers the potential to streamline workflows such as producing descriptive metadata, declassifying records, redacting PII, transcribing speech-to-text, or identifying objects within audiovisual content, among numerous other use cases.

While these workflows intersect with maintaining content authenticity and provenance (CAP) data, for the sake of limiting our scope, we are focused on a more holistic consideration of CAP. The LAMs community's main currency is trust, specifically the trust that the collections that are preserved and made accessible are what they say they are.[1]* That means protecting CAP data

[1] InterPARES Trust AI, "Authenticity." *Terminology Database* 2024. https://interparestrustai.org/terminology/term/authenticity. For the SAA's definition of authenticity, see the Glossary.
*Unless otherwise indicated, all online sites and resources were accessed February 7, 2026.

from accidental or malicious tampering, understood here as three critical stages: 1) generating metadata at the point of creation; 2) documenting a digital object's provenance, including any alterations generated by an AI tool; and 3) guaranteeing a digital object's authenticity at the point of access.

Emerging from the C2PA for G+LAM Community of Practice,[2] a Library of Congress-led initiative working with practitioners across the U.S., this report reflects the group's efforts to create an informal knowledge and learning network about C2PA and CAP concepts. In addition to this report, the group is gathering user stories to represent the needs of the government and LAMs community while returning feedback from the community to C2PA. Although the group's name as of this writing highlights C2PA, its expanded scope now accompanies CAP efforts beyond C2PA.

As a call-to-action to the preservation community, this document addresses a critical moment. AI technologies have captured the imagination of archivists, librarians, curators, and scholars in processing and accessing collections at scale. However, at the same time, essential CAP principles are at risk of being overlooked. We are witnessing in real-time the erosion of public trust in digital content distributed through news and social media as more examples of actual or suspected deepfakes appear daily. Thankfully, that skepticism has not yet encroached upon heritage repositories. But if left unchecked, AI technologies will have the same corrosive effects on LAMs' reputation established over centuries.

The research and perspectives represented within this report are a snapshot of the state of content authenticity and provenance as of early 2026, with an acknowledgement that within weeks and months of its release, parts of it may be outdated. The first two sections provide a brief primer on trust as LAMs' most essential currency and a history of CAP principles as applied to digital preservation. Following this overview is a selection of recent and ongoing research projects experimenting with how to secure CAP either using, or in response to, emerging AI technologies. The next section covers some of the overarching risks that AI poses for the LAMs community, culminating with a set of four "pillars" that serve as the foundation for a collective call-to-action.

Central to this call-to-action is the conviction that the solution will not, and should not, come from a single standard, institution, tool, or industry leader. Rather, it will require the collective effort of the community to conduct fundamental research and development, forge public and private sector collaborations, share results as widely as possible, and remain an active participant in the shaping of AI technologies driven primarily by industry interests. If there is one takeaway, it is that content authenticity must remain, as it always has, an endeavor that encompasses an assemblage of technologies, digital infrastructure, theory, standards, and most importantly, an ongoing series of complex decisions and partnerships among human practitioners.

---

[2] Kate Murray, Isabel Brador and Abbey Potter. "New Community of Practice for Exploring Content Provenance and Authenticity in the Age of AI." *The Signal*. Library of Congress. 2025. https://blogs.loc.gov/thesignal/2025/07/c2pa-glam/.

Trust as Currency

Trust is the essential currency of the LAMs sector. Universally, users expect that the collections stewarded by LAMs are secure from unintended alterations. While the presentation of the information may change, perhaps through a file format migration, digitization from an analog source to a digital file, or creation of an access copy for Web presentation, the expectation is that the intellectual content remains consistent over time.

In the LAMs community, provenance metadata captures the fundamental information of a source's creation and passage through time, documenting the various who's, what's, when's, where's and why's of its handling. It records the source's chain of custody, which enables users to build trust that the information they are interacting with is what it purports, and is documented, to be. A few examples of this in action include: MARC 21 Format for Bibliographic Data: 883,[3] which defines the provenance of metadata in data fields in the record, including if it was fully or in part machine-generated; Resource Description & Access (RDA) Metadata elements;[4] and Metadata Encoding & Transmission Standard (METS),[5] which includes structure for encoding descriptive, administrative, and structural metadata and many others.

In recent decades, digital technologies have complicated the verification of provenance data. Whereas a transaction of a physical object might involve recording the sale or loan from one owner or institution to another, the migration of a source from its physical instantiation to a digital one, or from one digital format to another, demands meticulous documentation of all technologies involved. Collection practitioners and users alike need to know which technology created the digital asset – for example, the make and model of the digital camera, the software version, or the use of plug-ins or scripts for editing – in addition to information about storage infrastructure such as the content or digital asset management system and server. They need the ability to declare with certainty that the content contained within an asset was authored by Person X using Software Tool Y on Date Z, and submitted to the repository via Workflow Q and stored on Digital Platform M.

But what happens when generative AI enters the mix? How do we account for an AI model as a potential creator or modifier? What aspects of its deep learning architecture must we record (if possible), from its training datasets to the unique properties of a particular model? How do we account for the fact that oftentimes an exact output is impossible to recreate? How do we document the extent to which generative AI has "remixed" original information, such as a work of art or piece of literature? What do we even mean by "original" and "authentic" content? Such

[3] Library of Congress. "Field 883: Metadata Provenance (R)," https://www.loc.gov/marc/bibliographic/bd510.html. May 2020.
[4] Library of Congress. *Resource Description and Access (RDA): Information and Resources in Preparation for RDA*. https://www.loc.gov/aba/rda.
[5] Library of Congress. *METS: Metadata Encoding & Transmission Standard. https://www.loc.gov/standards/mets/*. 2025.

questions must be addressed regardless of whether a digital object has come into contact with an AI system, as users will continue to expect the accuracy and verifiability of CAP data.

We want to note that trust in a source's authenticity does not extend to the veracity of the content contained within it – that is for researchers and content area experts to determine. Rather, the accelerating use of AI tools in content creation as well as digitization, description, and access workflows introduces new risks of misappropriation and misinterpretation.

What is clear is that the emergence of AI technologies, especially generative AI, has introduced an entirely new dimension of authentication of LAMs collection content for consideration. This moment demands deliberate, thoughtful, and sustained efforts to document and verify CAP data throughout the digital preservation lifecycle. As we shall see in the next section, that begins with acknowledging that the arrival of new technologies does not mean we should abandon our preservation principles, but rather adapt them for a new set of contingencies.

A Brief Historical Overview of CAP in Digital Preservation

Content authenticity and provenance are foundational to archival practice and records management, and are woven throughout archival literature, going back at least as far as the nineteenth century. The principle of *respect des fonds*[6] established that the authenticity of records derived from their documented relationships and chain of custody. No less an authority on archival practice than Hilary Jenkinson noted that authenticity is one of the distinguishing qualities of archives, and that "they are also by reason of their subsequent history equally free from the suspicion of having been tampered with in those interests."[7]

These foundational archival principles became especially critical – and complicated – with the emergence of electronic records in the late twentieth century.[8] Digital objects present unique preservation challenges: they are inherently mutable, technology-dependent, and may require interventions such as format migration or emulation to remain accessible and usable over time. The digital preservation community recognized that provenance and authenticity required mechanisms to adapt to these challenges. If digital objects might be transformed to mitigate technological obsolescence, how can institutions document these changes while maintaining trust in the objects' authenticity?

The Preservation Metadata: Implementation Strategies (PREMIS) data dictionary and XML schema are examples of standards developed by the digital library community to integrate

---

[6] Society of American Archivists. "*respect des fonds*." *Dictionary of Archives Terminology.* https://dictionary.archivists.org/entry/respect-des-fonds.html.
[7] Hilary Jenkinson. *A Manual of Archival Administration*. London: Percy Lund, Humphries & Co Ltd. 1937. pp. 12-13. https://dn720707.ca.archive.org/0/items/manualofarchivea00iljenk/manualofarchivea00iljenk.pdf.
[8] See, for example, Charles T. Cullen, et al. *Authenticity in a Digital Environment*. Washington, D.C.: Council on Library and Information Resources, 2000.

provenance and authenticity into digital preservation workflows.[9] Emerging from a working group coordinated by OCLC and Research Libraries Group from 2003 to 2005, PREMIS was introduced as a metadata schema specifically designed to support long-term digital preservation.[10] The data model supports provenance and authenticity by providing a structure to document relationships between objects and their origins, as well as changes over time, either due to format migration or other preservation events.

This era established a durable principle, namely that authenticity is inseparable from human judgment, policy, and stewardship, even when supported by technical controls. However, by the 2010s, as cultural heritage content increasingly circulated through aggregators, platforms, APIs, and social media, a structural limitation became clear: repository-centric provenance does not travel well. Metadata was frequently stripped, transformed, or divorced from content, while fixity information became irrelevant once files were copied or derivative versions were created. Provenance remained accurate within repositories, but became invisible downstream, precisely where questions of trust and authenticity were often most acute.

For cultural heritage institutions, the challenge was not only internal authenticity, but external interpretability. How do we signal stewardship, transformation, and responsibility once content has left controlled systems?

The National Digital Stewardship Alliance (NDSA) Levels of Digital Preservation, first published in 2013 and revised in 2019, introduced a simple and assessment-focused tool that provided targets to help build and evaluate the technological components of digital preservation programs, aimed especially at institutions developing digital preservation programs.[11]

The NDSA Levels are organized around five functional areas: Storage, Integrity, Control, Metadata, and Content. Content addresses essential components of authenticity along a progressively sophisticated range of approaches, from basic checksum generation and verification (Level 1) to fixity checking when moving or copying (Level 2), to maintaining audit trails of fixity checks and repairing corrupted files when necessary (Levels 3-4). These tiered recommendations acknowledge that most institutions have limited capacity for executing a preservation “gold standard.” Nevertheless, the NDSA Levels emphasize that integrity checking and documentation ought to form the foundation of trustworthy digital preservation.

These sample frameworks, grounded in archival traditions of provenance documentation, formalized through metadata schemas like PREMIS, and operationalized through implementation guidelines like the NDSA Levels, demonstrate how the preservation community has adapted long-held principles of authenticity and custody to the digital age. They recognize

---

[9] Library of Congress. *PREMIS: Preservation Metadata Maintenance Activity.* 2025. https://www.loc.gov/standards/premis/.

[10] For a brief history of PREMIS, see Micky Lindlar, Tracy Meehleib, and Karin Bredenberg. "PREMIS for All, for Good, Forever!", Open Preservation Foundation, 2022. https://openpreservation.org/blogs/premis-for-all-for-good-forever/.

[11] National Digital Stewardship Alliance. "NDSA Levels of Digital Preservation." 2026. https://ndsa.org/publications/levels-of-digital-preservation/.

that in digital preservation, authenticity is not only about maintaining bit-level stasis but about maintaining trustworthy documentation of an object's history, transformations, and context.

Understanding this historical pattern clarifies why current debates around the relationship between CAP and AI are not unprecedented, but an extension of long-standing preservation concerns within a new technical terrain. The central question has always been how to document change responsibly while sustaining trust. What has changed, as we shall see, is the scale, speed, and visibility of transformation, combined with communicating CAP beyond institutional walls.

Current AI Applications and Research in LAMs

In recent years, several novel approaches have experimented with securing CAP data in response to AI's impacts. The following examples are not intended as an exhaustive survey of methods, but rather are intended to emphasize the multiplicity of options and the need for additional collaborative research and development, which we address further in our concluding call-to-action.

Decentralization

As born-digital and digitized content increases exponentially, heritage institutions confront increasing challenges to maintain a central storage system that documents a digital object's full lifecycle, including its creation, chain of custody, alterations, format migrations, and other digital forensics metadata. Web3 technologies,[12] particularly blockchain, offer an opportunity to secure and maintain digital content integrity through a decentralized network of partnering institutions.

Project ARCHANGEL[13] tested the viability of blockchains as a decentralized ledger platform (DLT) for verifying archival provenance and authenticity with a two-year feasibility study (2017-2019) conducted in the UK by the University of Surrey, The National Archives, Open Data Institute, and University of Exeter. ARCHANGEL developed a distributed filesystem using a blockchain platform, Ethereum, which stored unique digital content signatures and the algorithms used to derive the signatures. That information, in addition to metadata such as an archivist's notes, versioning information, and content hashing were maintained by a network of participating LAMs. Any curatorial changes to the content were recorded and linked to all previous instantiations of the content, thus establishing the content's immutable digital provenance. Users could verify each instantiation of a digital object, which ensured full transparency of content modifications.

ARCHANGEL demonstrated the promise of using DLT to sustain long-term content integrity while exposing the resource-intensive hurdles inhibiting widespread adoption. As one proof of

[12] For an overview of current research and development in Web3 technologies, see web3 foundation. 2026. https://web3.foundation/.
[13] "ARCHANGEL - Trusted Archives of Digital Public Records." https://www.archangel.ac.uk/.

concept, the project team tested the blockchain platform on a small collection of videos to detect accidental or intentional file tampering that was invariant to the codec employed to encode the video. Audiovisual content was limited to 256 minutes to accommodate the block sizes of the blockchain platform. ARCHANGEL proved that while DLT implementation may be technically feasible, it is dependent upon cooperation among multiple institutions, ideally spread across wide geographic distances to ensure the platform's robustness. Each participating institution must possess appropriate staffing and infrastructure resources to install, maintain, and upgrade the system.[14]

If Project ARCHANGEL served as a proof-of-concept for decentralization, Starling Lab,[15] a partnership between Stanford University's Center for Blockchain Research and the USC Shoah Foundation, seeks to take Web3 technologies from concept to realizable workflows. Motivated by the immediate goal to secure the long-term preservation of the 56,000 oral testimonies of Holocaust survivors held by the Foundation, combined with the ambitious vision to expand to the wider historical, legal, and photojournalistic record, Starling Lab is developing frameworks that addresses content integrity at each stage of preservation: capture, storage, and verification. The frameworks leverage cryptography to encapsulate the digital asset at the point of creation to prevent further tampering at the bit-level. For now, the project team's approach applies C2PA to generate metadata such as the asset's timestamp, GPS coordinates, and the unique digital signature of the hardware used to create the asset such as a digital camera. The encrypted asset is then uploaded and stored in a decentralized system using blockchain registration reminiscent of the platform developed by ARCHANGEL. Throughout the process, Starling Lab applies what it calls "Authenticity-by-Design," which ensures that the preservation of provenance data is founded upon principles of "integrity, privacy, verifiability, persistence, and accountability."[16]

Model Context Protocol

Whereas Web3 technologies such as blockchain offer a decentralized approach to authenticating digital assets, Model Context Protocol (MCP) considers the issue from the opposite vantage point. Opening access to collection data to AI systems, particularly commercial generative AI systems seeking high quality training data, has raised concerns within the LAMs community. Beyond the questionable means by which content has been obtained and used for model training, current generative AI systems persistently distort, decontextualize, and outright fabricate information at a level unacceptable to the heritage community.

---

[14] Tu Bui, et al., "Archangel: Tamper-Proofing Video Archives Using Temporal Content Hashes on the Blockchain." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.* 2019.
[15] USC Shoah Foundation and Stanford Center for Blockchain Research. "starling lab." 2026. https://www.starlinglab.org/.
[16] Lindsay Walker. "The Starling Lab Framework." Filecoin Foundation for the Decentralized Web, 2024. https://ffdweb.org/blog/the-starling-lab-framework.

MCP “is an open-source standard for connecting AI applications to external systems.”[17] Functioning as a gateway between collection data and AI systems such as large language models, MCP can enable heritage institutions to retain control over access, thereby preserving content integrity. MCP servers can be developed to connect with collection metadata or the digital assets themselves. Users can compose natural language prompts within AI agents that instruct the agent to search and retrieve relevant data. Unlike standard search and discovery methods, AI agents introduce the possibility of retrieving ambiguous information that may not be formally expressed in a search term. In addition, AI agents offer the possibility of holding a ‘conversation’ with the database, asking follow-up questions that promote deep engagement with the collections, unearthing connections that may otherwise be difficult to surface.

MCP is still very much a protocol under development, with few LAMs as of early 2026 having developed working prototypes. MCP servers so far have been developed by individuals taking advantage of open datasets such as those offered by the Smithsonian Institution and the Rijksmuseum.[18] By following a few relatively simple steps, any user can install the relevant code typically found on a developer’s Github page and obtain a (free) API key from the institution. The server can be operated within frontier LLM environments (e.g., Claude Desktop) and accessed via natural language prompts typed within the standard text box.

One advantage of MCP servers is that LAMs have full control over what data can be accessed and with what “tools.” For example, the Rijksmuseum MCP server enables users to search artworks according to artist name, artwork type, period, materials, and other categories. In addition, users can retrieve information about a work’s historical context and physical properties, develop unique curated collections of available works, view high-resolution images, and create timelines that can track an artist’s career or artistic development.[19]

Because MCP servers control access to collection data, they also preserve content integrity. Prompts directed toward an MCP have a much higher likelihood, for example, of pulling accurate provenance information. With the Rijksmuseum, users can request information about a work’s exhibition history or any other ownership information recorded in its metadata. In addition, there is assurance that requests for copies of digital objects are authentic versions, since MCP servers are limited to extracting information maintained by the host institution. Each transaction occurs in real time, which means that any updates to an object’s metadata is instantly made available via the server, without further model training or changes to the code.

---

[17] Model Context Protocol. "What Is the Model Context Protocol (MCP)?" The Linux Foundation Projects. https://modelcontextprotocol.io/docs/getting-started/intro.

[18] As far as we can discern, the developers of the MCP servers used the museums’ open access datasets and were not formally endorsed by the institutions. Ruud Huijts. *Rijksmuseum MCP Server*. Github. https://github.com/r-huijts/rijksmuseum-mcp and Justin Molano. *Smithsonian Open Access MCP Server*. Github. https://github.com/molanojustin/smithsonian-mcp.

[19] Joshua Sternfeld. "Model Context Protocol and a Vision of Historical Research and Collections Access," *Encoding the Past*, 2026. https://encodingthepast.substack.com/p/model-context-protocol-and-a-vision.

Coalition for Content Provenance and Authenticity (C2PA) and Content Credentials

A recent approach gaining traction is the Coalition for Content Provenance and Authenticity (C2PA)[20] and its Content Credentials framework. At a high level, C2PA is the technical specification authoring body, sponsored by a large number of technology partners and part of a portfolio of "Digital Trust" projects within the Linux Foundation umbrella through the Joint Development Foundation, while Content Credentials are cryptographically signed data that provide verifiable information about a piece of digital contents' origin and editing history. There is increasing adoption of Content Credentials in technology-related industries. Examples of early or partial adoption include Google Pixel 10 phones,[21] Ring doorbell cameras,[22] support in various Adobe products,[23] Sony Camera Authenticity Solution,[24] and many more.

In practicality, C2PA is an extension of existing content authenticity and provenance practices, not a replacement. Its real value to the LAMs community is its purpose-built capacity to document interactions of content with AI at any point in the lifecycle as well as verification through an external trust center. This interaction is particularly valuable at points of dissemination and reuse. C2PA can carry forward a verifiable statement that a digital object was produced, transformed, or released under the stewardship of a trusted repository. For example, a repository could use C2PA action assertions to indicate that a file is an access derivative, that AI-assisted processing occurred under documented policies, or that integrity checks were performed. In short, C2PA does not prevent the alteration of original source materials by generative AI, but in situations where knowing a digital object's provenance is critical, it can provide a layer of authentication.

Challenges for LAMs

The three experimental approaches to addressing content authenticity and provenance suggest that significant challenges for LAMs remain. Implementing CAP processes amidst AI developments will place demands on capacity for institutions that are already stretched thin. LAMs will face difficult questions about how to allocate limited resources for technologies that often seem to be in a perpetual state of experimentation. Yet, not doing anything poses its own risks. AI technologies, particularly generative and autonomous systems, threaten LAMs' capacity to maintain content authenticity not in some distant future but right now, and their ubiquity will only increase over time.

---

[20] C2PA. "Content Credentials: C2PA Technical Specification." https://spec.c2pa.org/specifications/specifications/2.3/specs/C2PA_Specification.html.
[21] Eric Lynch. "How Pixel and Android Are Bringing a New Level of Trust to Your Images with C2PA Content Credentials." *Google Security Blog*. 2025. https://security.googleblog.com/2025/09/pixel-android-trusted-images-c2pa-content-credentials.html.
[22] Ring. "Your Ring Videos, Verified." *Ring*. 2026. https://blog.ring.com/about-ring/your-ring-videos-verified/.
[23] Adobe. "Adobe Content Authenticity." https://contentauthenticity.adobe.com/.
[24] Sony. "Camera Authenticity Solution." https://authenticity.sony.net/camera/en-us/.

We have identified three interrelated challenges with direct implications for sustaining content authenticity and provenance that administrators must confront regardless of other emerging standards, tools, or workflows. First, as with adoption of earlier digital preservation practices, pivoting to address the impact of AI will introduce infrastructural demands requiring careful deliberation and planning. Second, ethical and privacy concerns that accompany AI technologies will reorient how we think about which CAP data needs to be accessible and why. Third, AI developments will continue apace, with or without the LAMs community's input. Processes that are viable today may be outdated within years, if not months, potentially negating progress toward retooling institutional capacity.

Infrastructural Demands

The era of digital preservation required a massive and costly transition away from analog methods of processing and storage. While the benefits of scale and enhanced access were clear, there were trade-offs that included the vulnerability of hardware and software to failure; equipment obsolescence; the need for education and training, from higher education programs to mid-career professionals; and extensibility across competing systems and intellectual frameworks.

Artificial intelligence presents the potential for even deeper disruption along these same tracks. Workflows and job tasks will undoubtedly need to adapt to the changes that AI brings. This will have widespread trickle-down impacts on training and skills development infrastructure as AI brings requirements for entirely new expectations for task-based knowledge.

CAP standards and applications may be complicated and labor-intensive to implement and maintain. Any adjustments to existing workflows must justify the time and resources invested in them, including human-centered work. LAMs will need to consider if the return on investment is in line with wider institutional goals including how the data would be gathered, stored, and used over time.

Adoption of new CAP standards may incur technical debt to maintain the infrastructure needed for trust centers, data integrity, and migration. CAP will require a buildup of resources over time that will need to be maintained. But equally, non-adoption or delayed engagement with CAP may result in lack of integration and compatibility with structured data from CAP-aware platforms, vendors, and infrastructure providers.

There's also a risk of a dependency knowledge debt in which LAMs may be relying on external vendors and platforms to apply or manage CAP data without expert internal understanding of the process, resulting in a loss of control over verification, migration, trust assurance and long-term accessibility and access of collections content.

LAMs thus need to consider the risks of conceptual technical debt which could build up if they do not engage with the standards and specifications around CAP data such as C2PA, which might force institutions with rich, nuanced archival practices into ill-fitting technical constraints determined by industry, not cultural heritage, needs. Leadership engagement today to help

shape the standards and expectations can be a form of knowledge and technical debt prevention.

Ethics and Privacy

For some communities and constituents, there are mature and valid concerns about privacy risks with CAP data presented by AI. CAP data is purposely a revealing process which may expose sensitive information that may be harmful to some groups. The same detailed level of data necessary to record the creation of a digital asset while avoiding manipulation by AI systems, including the identity of the creator, timestamps, the unique signature from the equipment, and the geographic coordinates of the site of creation, can have harmful effects once they travel far beyond institutional context and are used for unforeseen purposes. This can raise concerns about personal safety, institutional security, and exposure of sensitive operational details.

In addition, CAP data might pose collection-specific risks for individuals or communities, especially those that have been historically marginalized or underrepresented. Provenance data that records AI-assisted enhancement, reconstruction, or interpretation could inadvertently surface information that conflicts with ethical commitments, donor agreements, or community protocols. Even when the content itself is public, the documentation of how it was processed, by whom, using which tools, and under what assumptions, may introduce new vectors for harm, misinterpretation, or unwanted scrutiny. Because CAP data is designed to be machine-readable and interoperable, these risks can be amplified at unprecedented scales with AI technologies.

There is a long-term privacy risk tied to persistence and aggregation. Long-lived provenance records can become problematic as social norms, legal frameworks, or institutional practices change. What seems like appropriate transparency today may be considered excessive disclosure in the future, especially if provenance data is aggregated across collections. It is difficult to predict how advanced algorithmic capabilities of AI systems might deliberately or inadvertently infer patterns about staff behavior, institutional priorities, or sensitive workflows. The risks could be akin to reconstructing an individual's profile simply from collating anonymous data gathered from a cell phone's location over time.

For this reason, leadership engagement is critical to ensure that CAP data is paired with clear governance: defining what information is appropriate to assert, what should remain internal, how roles are abstracted, and how privacy, ethics, and care obligations are balanced against transparency. One of the central risks is that otherwise well-intentioned institutions adopt processes without embedding privacy-aware decision-making into their provenance strategy from the outset.

CAP implementations can introduce risks that are especially acute for cultural heritage institutions, where context, interpretation, and ethical judgment are central to stewardship by excluding humans from the loop in content authentication and provenance processes. Fully automated systems can reliably attest to technical facts, such as whether a file has changed or whether a cryptographic signature remains valid, but they cannot assess meaning, intent, or appropriateness. Without human review, there is a risk that automated provenance signals will

be treated as authoritative endorsements rather than as structured disclosures, leading users to place too much trust in AI-processed or algorithmically generated representations of cultural heritage materials. This is particularly dangerous when AI tools introduce subtle distortions, bias, or speculative reconstruction that only a knowledgeable professional can recognize and contextualize.

There are also risks of ethical blind spots and harm amplification when humans are removed from decision points. Automated pipelines may apply authentication or provenance assertions uniformly, without regard for culturally sensitive materials, donor restrictions, or community protocols. In such cases, machine-generated provenance could legitimize transformations or disclosures that a human curator, archivist, or community liaison would have flagged as inappropriate. Over time, this can erode institutional commitments to care, consent, and respect, especially for collections involving living communities or contested histories. The absence of human oversight makes it harder to detect systemic bias or errors embedded in AI tools themselves, allowing those issues to propagate silently across large collections.

Speed of Development of AI Technologies

Implicit throughout the two challenges addressed above is the breakneck rate at which AI technologies are advancing. An institution may have well-intentioned goals to improve institutional capacity by upgrading asset management platforms and storage infrastructure, encouraging staff to obtain additional training by earning online credentials or certifications, or even hiring new staff conversant in AI technologies.

While these measures and others may be necessary to stay apace, they may still prove insufficient. Not only will the cycle of updates and upgrades likely accelerate, but there are technological shifts that will test practitioners in new ways. For example, AI systems are already redefining how researchers in numerous fields, including the sciences, engineering, and social sciences conduct their work. For institutional repositories charged with preserving their work, how do they account for the presence of AI agents conducting computational work exceeding human levels of productivity? Relatedly, as AI enables everyone, including those without coding expertise, to build AI agents that can streamline work, process information, or conduct complex analysis, which elements ought to be preserved? What does the chain of custody look like for a project conceived of by humans, yet predominantly executed by (proprietary) AI systems?

We are on the cusp of an explosion of automated workflows, generated outputs, software, and metadata that will test the preservation community in our ability to keep abreast of swift developments. In some respects, the work may be familiar, yet at a larger scale. In other respects, they may demand new approaches, methodologies, or standards. Regardless, we must continue to apply the values of preserving authenticity and provenance that have served us for centuries.

Call-to-Action

In the eyes of creators, users, and stewards, it's possible that there has never been a more important time to devise a strategy to preserve the authenticity and provenance of digitized and born-digital content throughout its lifecycle. From this standpoint, LAMs are in a unique position. Due to their long-established role as stewards of trustworthy information sources, their influence and advocacy have the potential to traverse communities. By purposefully collaborating with those upstream of their positions, LAMs can impart the values and strategies for securing trust in content authenticity and provenance data. LAMs should serve as the steadfast voice in the face of transformation while simultaneously advocating for reasonable adaptation.

With this in mind, this document proposes a set of four pillars that can guide us through turbulent times. Like the values of authenticity and provenance themselves, these proposals ought to feel familiar. While the practices themselves may radically change, the path toward productive change should not. This is not a moment for the preservation community to erect silos, which will prove unsustainable given the rate of technological change. As the examples below demonstrate, there is an openness to unconventional collaborative arrangements that traverse geographic, organizational, and professional boundaries. The community must overall advocate not only for practical innovation, but how to communicate those breakthroughs.

Research and Development

The LAMs community functions best with shared standards, workflows, tools, and practices from which to conduct their work. AI technologies demand complex decisions whether to adopt existing authenticity and provenance methods or develop new methods whole cloth to accommodate new realities. Instinctively, practitioners understand that humans must remain "in the loop" overseeing decision-making along every step, but striking the appropriate balance between oversight and technological autonomy remains an open set of questions.

Research and development is necessary to interrogate each stage of authentication and generative provenance metadata. For it to be effective, R&D work should balance the theoretical aspects of CAP that draws upon and updates CAP principles while strategically applying them to numerous AI scenarios. From basic research and one-off case studies to sustained multi-institutional applied research, this work must assure representation from the diverse preservation community, ranging from small historical societies and local archives to leading museums and libraries.

Fortunately, we do not need to start from square one. Beyond the projects mentioned earlier that are experimenting with different approaches to securing CAP, there has been considerable activity in both theoretical advancement and tool development. A multi-national group of researchers known as InterPARES Trust AI, or ITRUSTAI, has undertaken an expansive set of research studies "aiming to design, develop, and leverage Artificial Intelligence to support the

ongoing availability and accessibility of trustworthy public records."[25] The five-year project (2021-2026) has yielded considerable research output in the form of case studies, analysis of specific AI tools, and establishment of archival principles that can inform development of AI preservation technologies. Beyond their body of work, which is too expansive to cover here, ITRUSTAI has cultivated a community of researchers dedicated to advancing an evolving understanding of CAP. Over time, their research will spur new areas of inquiry, as well as undoubtedly filter down into MLIS curriculum development.

We should also not overlook the considerable R&D activity conducted beyond the field of preservation. For example, researchers in industry and academia continue to experiment with watermarking as a method of conferring authenticity by embedding "AI fingerprints" within AI-generated content, even large language models.[26] A similar race toward identifying and evading deepfakes has been underway for several years, with direct implications for CAP.[27] While both areas have yielded promising results, they have not yet risen to a level that would meet LAMs' high standard of trust.

Partnerships and Collaborations

One of the great legacies of the LAMs community is its successful reliance on collaboration. None of this work happens in a vacuum. It's important that the community come together to support aligned efforts to avoid duplication of work and to further standards and good practice-based work. This report has explored some longstanding work, such as the NDSA Levels of Digital Preservation, but the AI challenge has brought new players to the field. These include the Digital Object Authenticity Working Group (DOAWG),[28] "an independent, voluntarily working group established in the beginning of 2024 by interdisciplinary professionals within the GLAM community (Galleries, Libraries Archives and Museums) which is working toward a solution for addressing the problem of authenticating cultural heritage images."

Another is Trust in Archives (TAI),[29] a "cross-disciplinary coalition dedicated to preserving the authenticity of media collections, encouraging transparency, and ensuring public trust in

[25] Interpares Trust AI. https://interparestrustai.org. The international project team is organized by working groups addressing various archival principles and practices such as Appraisal and Acquisition, Arrangement and Description, Retention and Preservation, and Creation and Use. To date, they have released nearly 50 studies. They describe their work as "aiming to design, develop, and leverage Artificial Intelligence to support the ongoing availability and accessibility of trustworthy public records by forming a sustainable, ongoing partnership producing original research, training students and other highly qualified personnel, and generating a virtuous circle between academia, archival institutions, government records professionals, and industry…."

[26] See, for example, Siddarth Srinivasan. *Detecting AI Fingerprints: A Guide to Watermarking and Beyond*. Brookings Institution. 2024. https://www.brookings.edu/articles/detecting-ai-fingerprints-a-guide-to-watermarking-and-beyond/ and Aloni Cohen, Alexander Hoover, and Gabe Schoenbach. *Watermarking Language Models for Many Adaptive Users*. In *2025 IEEE Symposium on Security and Privacy (SP)* (pp. 2583-2601). IEEE. 2025.

[27] The following project from MIT Media Lab's Affective Computing Group provides an overview of recent research into deepfake detection: Matt Groh. "Detect Deepfakes: How to Counteract Misinformation Created by AI." MIT Media Lab. 2025. https://www.media.mit.edu/projects/detect-fakes/overview/.

[28] Digital Object Authenticity Working Group. 2026. https://www.doawg.org/.

[29] Trust in Archives Initiative. 2026. https://www.trustarchives.org/.

archives. In an era increasingly influenced by synthetic media, TAI brings together expertise across sectors to develop practical tools and resources that address the challenges of generative AI, ensuring archives remain trusted sources of the historical record." TAI has several components including an Authentication Working Group,[30] which aims to "to help archives of all sizes evaluate available options and implement the most effective methods of authentication and attestation for their unique contexts, ensuring that collections remain trusted sources of the historical record."

One more potential partnership to explore is the AI and Multimedia Authenticity Standards Collaboration,[31] which is "a global, multistakeholder initiative led by the World Standards Cooperation (IEC, ISO, ITU)...[and] brings together global leaders in standards, technology, academia and policy to support the journey toward building trust in digital media."

The important takeaway is that this is collaborative work, something that the LAMs community excels in. These changes in workflows and technologies and staff roles will impact all corners of digital preservation across many sectors in the years to come. The key is to partner together, to share ideas, concerns, solutions and challenges to maintain our history of transparency, but also to build on economies of scale, resources, and attention.

Advocacy with Industry, Vendor and User Communities

Within the digital realm, LAMs have widely embraced practices and systems that leverage open methods for computationally verifying the integrity of digital objects. LAM-led practices augment IT industry and vendor practices, enabling a higher visibility into process, independent validation of results, and vendor-agnostic implementations. These are essential for long-term preservation of digital collections.

Key to this work is advocacy about the imperative of CAP data for LAMs communities of practice, their users and with industry and vendor communities who drive both standards and tool development. As repositories of diverse data, LAMs are in a unique position to take leadership in the development and implementation of CAP practices that augment industry-specific practices that develop around area-specific features.

Despite a variety of different institution types under the LAMs category, securing digital collection items' CAP data requires standardized, vendor-agnostic, and open specifications to achieve base goals and ensure institutional trustworthiness in the face of changes across the user-facing technology landscape. LAMs are uniquely suited to join existing industry-led groups in developing open and shared specifications that address LAMs use cases and allow production and validation utilizing a public-private ecosystem of tools and implementations.

[30] Trust in Archives Initiative. "Authentication." 2026. https://www.trustarchives.org/authentication/.
[31] World Standards Cooperation. "AI and Multimedia Authenticity Standards Collaboration." https://aiforgood.itu.int/multimedia-authenticity/.

One example is active participation in standards bodies such as ISO 171/TC 2/SC 13: Content Provenance,[32] which is the working group formalizing the C2PA Content Credentials specification for international standardization. Codifying C2PA as an international ISO standard can increase adoption, support products with reproducible and consistent quality, and encourage more system and tool development. All of these mean more options for LAMs workflows.

Another option is providing feedback to industry and vendor partners. Once LAMs communities identify their CAP needs, especially those impacted by AI, we can establish expectations. One example of this is the C2PA for G+LAM User Stories which were shared with C2PA specification authors to highlight the specific needs of the digital preservation community.[33] This collaborative group also shared feedback on the Content Credentials specification which were incorporated into the standard in a subsequent version. While C2PA Content Credentials were not specifically aimed at digital preservation and LAMs needs, the LAMs community can influence their continued development to then take advantage of the subsequent momentum and tools that go along with this industry driven resource. We don't have to start from scratch but we also need to tell our allied and aligned communities what we need.

Open Distribution of Results, Lessons Learned, Ideas

The LAMs community must innovate in communicating results, resisting a tendency to maintain siloed activity. The traditional path toward establishing field-wide standards and practices, which can take years of deliberate planning and iterative applied research, in certain scenarios may not be sustainable in a culture where technological innovation is occurring on a more rapid time-scale. Withholding project results until they are peer reviewed and published within a scholarly publication, which can take over a year, will likely be outdated by the time they are publicly accessible.

That is not to say that the community should abandon traditional scholarly communication channels, which will continue to serve as a vital outlet for formalizing progress. Academic publications, edited anthologies, efforts coordinated by professional associations, and other established outlets must continue to validate principles and approaches that can be revised or adapted according to shifts in AI developments.

In the meantime, the LAMs community should share experimental findings as they unfold. This demands exploiting the tools of real-time communication, including social media, blog posts, Github, and others. Taking a page out of the science community, practitioners should not shy away from publishing pre-print materials on sites such as arXiv.

---

[32] For more information about the ISO standard and the activities of its working group, see International Organization for Standardization. "ISO/TC 171/SC 2: Document File Formats, EDMS Systems and Authenticity of Information." https://www.iso.org/committee/53674.html.
[33] Murray, Brador and Potter. "New Community of Practice for Exploring Content Provenance and Authenticity in the Age of AI."

In short, the moment demands a more radical approach to open communication. To be clear, we are not advocating solely for an endorsement of AI-enabled progress. The challenges to maintaining institutional trust and sustainable CAP data necessitate ongoing deliberation. LAMs practitioners should continue to question whether new methods, tools, and practices uphold principles of content authenticity and provenance. Innovations will invariably raise unforeseen consequences. As the community encounters roadblocks and dead ends, there should be a willingness to share lessons learned. Failures are just as valuable as breakthroughs.

To facilitate such enhanced real-time communication, the LAMs community needs to coalesce and create new spaces for sharing experiences and results. The international body AI4LAM[34] offers one model, facilitating a combination of workshops, working groups, and annual conferences, along with online communication channels such as Slack, a Google Group, and YouTube channel. This approach encourages communication at multiple levels, from international collaboration to regular meet-ups for regional chapters, which currently include the Washington, D.C. metropolitan region, Europe, and Australasia.

Institutions needn't join formal organizational bodies to reap the benefits of continued dialogue. A constellation of local institutions might decide to regularly convene to share knowledge and insights. Such informal gatherings hold certain advantages over larger networks, because they can address issues pertinent to shared local collections and their CAP data that otherwise might go unaddressed in wider forums. Practitioners may be more willing to share their experiences and ask questions in a local context versus formal networks.

Conclusion

This report arose from a simple question that all LAMs administrators and practitioners have and will continue to face: Amidst dizzying AI developments, what should my organization do? A societal transition is underway that is destabilizing our relationship to digital content. This shift will have profound effects on the LAMs community's reputation as trusted repositories. The purpose of this report was not to offer concrete tools or practices, but rather to provide a combination of theoretical grounding, an overview of current work, and a set of action-oriented pillars that might assist in decision-making.

Institutional capacity and mission priorities might inhibit participation in the efforts identified above. What is clear, however, is that this is not a moment to sit idly by. At a minimum, LAMs professionals should familiarize themselves with the basics of emerging technologies. Reading a handful of articles or blog posts, reaching out to colleagues to ask questions, attending a workshop, or engaging in dialogue will clarify institutionally-specific CAP needs.

[34] AI4LAM. 2026. https://sites.google.com/view/ai4lam.

Engaging with the wider LAMs community in actions large and small will demystify AI. As disorienting as these technologies may seem in challenging our understanding of truth and authenticity, taking a proactive stance will reinforce technologically-agnostic shared knowledge. Not every institution will be in a position to adopt AI technologies, nor should they. However, all LAMs ought to lead in the defense of CAP principles, which will not only sustain their role as heritage stewards but offer a critical perspective that can resonate across sectors.

Glossary

Artificial intelligence or AI: A sub-field of computer science research and practice that aims to develop computational models to simulate or approximate human cognition, behaviour, decision-making, and reasoning.[35]

Authenticity: The quality of being genuine, not a counterfeit, and free from tampering, and is typically inferred from internal and external evidence, including its physical characteristics, structure, content, and context.[36]

CAP: Content authenticity and provenance

Generative AI or GenAI: The class of AI models that emulate the structure and characteristics of input data in order to generate derived synthetic content. This can include images, videos, audio, text, and other digital content.[37]

LAM: Libraries, archives and museums

Provenance: Information regarding the origins, custody, and ownership of an item or collection.[38]

[35] InterPARES Trust AI, "Artificial Intelligence." *Terminology Database* 2024. https://interparestrustai.org/terminology/term/artificial%20intelligence/en.
[36] Society of American Archivists. "Authenticity." *Dictionary of Archives Terminology.* https://dictionary.archivists.org/entry/authenticity.html.
[37] National Institute of Standards and Technology. "Generative Artificial Intelligence." *Glossary.* *https://csrc.nist.gov/glossary/term/generative_artificial_intelligence*.
[38] Society of American Archivists. "Provenance." *Dictionary of Archives Terminology.* https://dictionary.archivists.org/entry/provenance.html.

About the Authors and Contributors

Authors

Kate Murray - Kate Murray is a Digital Projects Coordinator at the Library of Congress where she leads the Federal Agencies Digital Guidelines Initiative (FADGI) Audio-Visual Working Group and the Sustainability of Digital Formats website. Kate is a member of AMIA (Preservation Committee co-chair 2010-2013), IASA (Technical Committee member), SMPTE and ISO/TC171/SC2 (PDF standards committees, including WG13 Content Provenance). She is the co-founder of the C2PA for G+LAM (Coalition for Content Provenance and Authenticity for Government plus Libraries, Archives and Museums) Community of Practice in 2025. In 2019, she received one of the inaugural Joint Technical Symposium (JTS) Awards for outstanding contributions to the technology of the audiovisual archiving field. In 2021, she received the NDSA Excellent Award for Individuals for making a significant contribution to the digital preservation community through advances in theory or practice. In 2022, she was a finalist for the DPC 20th Anniversary Award on behalf of FADGI's cumulative body of work. A global leader in digital file formats research and audiovisual preservation, she is the LC liaison to the PDF Association, Vice Chair of the Digital Preservation Coalition's (DPC) Executive Board and Vice Chair of the DPC Americas stakeholder committee. Kate received her undergraduate degree in Medieval Literature from Columbia University and her MBIBL from the University of Cape Town.

Contact: **kmur@loc.gov**

Joshua Sternfeld - Joshua Sternfeld is a historian, program leader, and heritage strategist whose work focuses on the implications of artificial intelligence for libraries, archives, and museums. He spent more than fifteen years at the National Endowment for the Humanities (2009-2025), serving as Senior Program Officer and Assistant Director in the Division of Preservation and Access, where he coordinated national initiatives supporting archival stewardship, digital preservation, and research and development in emerging technologies. His work included managing programs that enabled LAMs to experiment with machine learning, computational analysis, and data-driven approaches to large-scale cultural collections, as well as helping to launch and shape the *Humanities Perspectives on AI* initiative to advance field-wide conversations about ethics, interpretation, access, and long-term stewardship. He worked closely with partners such as the Library of Congress, the National Gallery of Art, and public media organizations to strengthen digital infrastructure and support responsible innovation across the cultural heritage ecosystem. In parallel with his program leadership, Sternfeld is an active scholar whose writing on digital history, LAMs, and AI has appeared in venues such as the *American Historical Review*, *American Archivist*, and *Transactions of the American Philosophical Society*. He also writes for a broader professional audience through his Substack, *Encoding the Past*, and currently works as an independent consultant and advisor on AI strategy, research integrity, program development, and responsible innovation.

Contact: **joshsternfeld@gmail.com**

Contributors

David Cirella - David Cirella is the Head of Digital Preservation at Yale University, where he specializes in the preservation of digital content and the intersection of library collections, technology, and the needs of users.

Ann Hanlon - Ann Hanlon is Head of Digital Collections & Initiatives at the University of Wisconsin-Milwaukee Libraries, where she has led digital collections, preservation, and digital humanities services since 2012.

Nick Krabbenhoeft - Nick Krabbenhoeft is Assistant Director of Digital Preservation at the New York Public Library, where he has led the implementation of core software, built collaborative teams, and guided workflow improvement from documentation initiatives to petabyte-scale data transfers.

Eric Lopatin - Eric Lopatin is Product Manager for Digital Preservation at the California Digital Library (CDL), where he leads the product development of CDL's preservation repository and directly supports University of California campus library and systemwide digital preservation initiatives.