LIBRARY OF CONGRESS

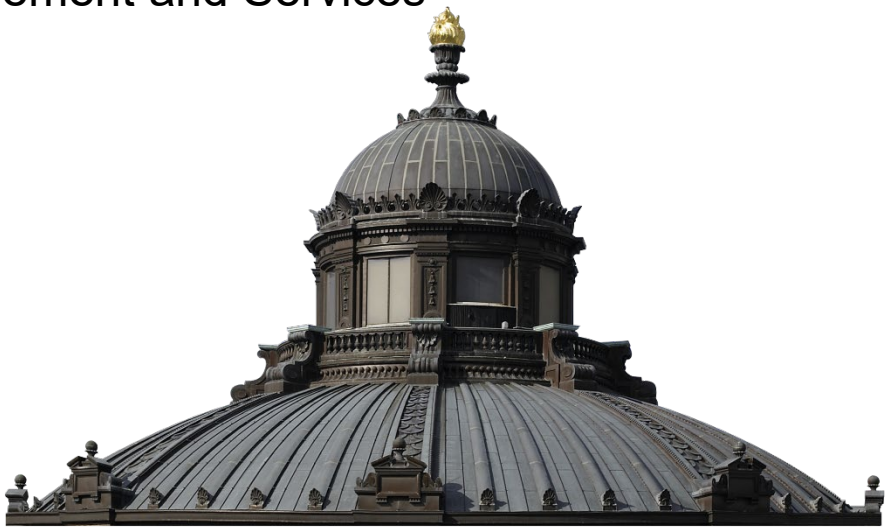
March 24, 2025

C2PA for G+LAM

Coalition for Content Provenance and Authenticity for Government and Libraries, Archives and Museums

Kate Murray, Digital Collections Management and Services
Abbey Potter, LC Labs

Designing Storage Architectures (DSA)



What is C2PA?

C2PA: Coalition for Content Provenance and Authenticity (https://c2pa.org/)

The C2PA standard defines a set of metadata containing details about the provenance of information displayed on a digital device including, for example, images, video, sound or text data both as standalone files and embedded within files.

Similar to a nutrition label, the C2PA metadata for this information can include, among other things, the publisher of the information, the device used to record the information, the location and time of the recording or editing steps that altered the information. The C2PA metadata as well as the main content is secured with hash codes/fixity values and certified digital signatures.

How does C2PA work?

- **Provenance data:** C2PA uses statements called assertions to describe the history of a piece of content, including who created it, when and where it was created, and how it was edited.
- **Cryptographic hashes**: C2PA uses cryptographic hashes to bind to each pixel in a source image or video.
- **Digital signing**: C2PA uses digital signatures, such as X.509 certificates, to make the content tamper-evident.
- **Content Credentials:** Content Credentials implement C2PA's technical standard and cryptographic methods to protect information from being tampered with.
- **C2PA Manifest:** A C2PA Manifest is a verifiable unit that binds together assertions, claims, credentials, and signatures.
- **C2PA Manifest Store**: The C2PA Manifest Store stores the set of C2PA Manifests that represent the provenance data of an asset.

Standardization of C2PA

The C2PA specification is fully published at https://c2pa.org/specifications/specifications/1.3/specs/C2PA_Specification.html with issues tracked on https://github.com/c2pa-org/specifications/issues.

The spec is now under fast track to become an ISO specification under TC 171/SC 2/WG 13 Content Provenance as ISO 32008 to “Authenticity of information - Extensions to Content Credentials for ISO 32000-2 (PDF 2.0)”.

Also JPEG Trust - implementation of C2PA for JPEG: ISO/IEC PRF 21617-1 (ISO/IEC JTC 1/SC 29): https://jpeg.org/items/20241202_press.html

How are C2PA/Content Credentials stored?

Content Credentials can be stored and recovered in three ways: (from: https://helpx.adobe.com/creative-cloud/help/content-credentials.html)

- Attached directly to their respective files
- Published to the Content Credentials cloud
- Attached and published

C2PA and responsible AI

Content provenance securely and cryptographically attaches information (metadata), known as Content Credentials, to photos, videos, or audio so consumers can understand where it came from, how it was created, and whether it has been edited.

For example, Content Credentials can show if something was made by a camera or generated by AI, when and where it was created, and any changes it has gone through. This helps people understand what they see online by providing a clear record of its origins and history.

While not infallible, C2PA is a tool towards supporting responsible AI.

Risks and weaknesses

- C2PA may be exposed to domains, Example use case from CBC / Radio Canada https://www.hackerfactor.com/blog/index.php?/archives/P3.html
- Witness: Tomorrow's Great Digital Divide: Content With or Without Provenance: https://blog.witness.org/2025/03/tomorrows-great-digital-divide/
- Lots of enthusiasm now but what will actual adoption look like in the coming years?
- Summary is that C2PA is still an emerging practice but there is much enthusiasm across multiple and diverse domains so it is worth exploring for the G+LAM communities

C2PA for G+LAM

- LC AI Working Group briefing in December 2024
- LC and friends initiative started in January 2025
- Goals:
  - Awareness building about C2PA for the government and LAM communities
  - Create informal community of practice/discussion about C2PA concepts and potential implementation
  - Collaborate on use cases, aside from Records Management, for the government and LAM communities
  - Return feedback to C2PA
- Project details and notes: http://bit.ly/4ioaNQa

Selected draft use case ideas

- An institution wants to document stewardship of specific instances of content, especially for digital collections that are often shared and duplicated and sometimes altered.
- Embedding C2PA Manifest into Broadcast Wave files as part of potential update to Embedding Metadata in Broadcast WAVE Files - Federal Agencies Digital Guidelines Initiative (FADGI)
- How can C2PA be leveraged in the Acquisition process if incoming content already has C2PA?
- Can C2PA reduce paperwork/improve efficiency for requests of true and accurate “certified copies”?
- What role can C2PA play for using historical maps to prove land claims for both accuracy of image but also associated metadata?

Interested in participating (if G+LAM)?

- Use case subgroup meeting is 4/4/2025, 1-2:30pm – small group working meeting
- Community quarterly calls planned for the first Friday of every 3rd month. So June 6, Sept 5, and Dec 5 in 2025. More details to come.
- Contacts:
  - **Abbey Potter**, Senior Innovation Specialist, LC Labs: abpo@loc.gov, LinkedIn: https://www.linkedin.com/in/abigailpotter/
  - **Kate Murray**, Digital Projects Coordinator, Digital Collections Management and Services: kmur@loc.gov, LinkedIn: https://www.linkedin.com/in/kate-murray-581a982/

Thank you and questions