12/22/25
D. Zellner, Q. Sluzenski, K. Warner

2025 Content Credentials Progress Report

The Need

Provenance, the origin of materials and their contents, has always been a concern for libraries. How a digital file was created and its origin is equally important, especially with the advent of AI.

Citation builds on provenance; a more sophisticated way of citing a digital source provides the user with more context in an immediate way.

Content Credentials are a method to address these needs so that image, audio, and video files contain provenance information that users can see.

Content Credentials - Definition

Adobe, a major participant in the Content Authenticity Initiative (CAI), defines Content Credentials as

> *"a durable, industry-standard metadata type that acts like a digital nutrition label for content. These credentials record information about a work's creator and production history, including whether it was captured by a camera, generated by AI, or edited using tools such as Photoshop."*

Technically, Content Credentials consist of a set of JSON-based manifests stored in a "manifest store." These manifests are generated at various points in a digital object's lifecycle and follow a standard developed by the Coalition for Content Provenance and Authenticity (C2PA).

The cultural heritage community has approached C2PA and its advocacy organization, CAI, with caution due to their close association with Adobe. Although the tools are currently open source, there are concerns that users may ultimately be steered toward proprietary solutions. In addition, cultural heritage institutions are less prominently represented than industries such as advertising, news, and commerce, resulting in solutions that are more closely aligned with those sectors' needs.

We believe that it's in Northwestern's interests to pursue using C2PA because the infrastructure is in place, and there is critical mass behind this solution. These facts outweigh the previously stated concerns.

The Life Cycle of a Digital Object

1. Pre-Ingest into the Digital Repository

Tools that create and process digital files generate technical metadata either embedded in the file or as a side car file. These metadata files can be considered manifests that can be included in the manifest store (Content Credentials) — most are *not* JSON files.

2. Post-Ingest into the Digital Repository

Currently our repository tools do not embed or create side car files that record the actions taken on the files. The server delivery of files, including streaming media and image delivery via the IIIF server, further complicates the objective of having provenance information accompany the files.

3. Signing the Digital Files

Signing is a critical component of Content Credentials. Much like a bookplate in a library book, a digital signature establishes the origin of a media file. Through this signature, users can verify a file's provenance, supported by trusted lists that are created and maintained by third parties. How often files must be signed has not yet been determined; while frequent signing could establish a detailed chain of custody, it may not be necessary for all workflows.

Content Credentials and the User

There are two methods for delivering Content Credentials to users:

Hard binding: The user downloads the file with the Content Credentials embedded directly in it.

Soft binding: The Content Credentials are referenced externally, allowing the user to access them via a link (for example, in an IIIF manifest). This approach assumes the existence of a repository of Content Credentials. In Northwestern's case, these credentials would be stored in the institutional repository as associated data. This model also allows staff to update Content Credentials over time to reflect changes to the physical or digital item.

Content Credentials Research at Northwestern

Following Dan's attendance at the Content Authenticity Summit in New York, he partnered with Quinn Sluzenski to research Content Credentials, with Kayla Warner joining as a liaison to Library IT. Together, they met with stakeholders across the

library—including curatorial, metadata, IT, marketing, and other groups—to present the concept and gather feedback. They also created mock-ups and began defining the provenance metadata that users would see. Below are notes gathered from this work.

Digitization (Dan Zellner)

Prior to ingest, capture devices and software generate a substantial amount of metadata. During quality control, additional data is created that documents file quality and is useful both to users and to digitization staff at other institutions. Digitization vendors also supply technical metadata that must be incorporated into Content Credentials. It will be necessary to determine how this information is captured and managed, and whether this will require additional vendor requirements in statements of work. Provenance will be especially important as the Big Ten Academic Association moves toward the "Big Collection," a shared compilation of holdings from Big Ten member institutions.

Descriptive Metadata (Quinn Sluzenski)

The goal of the descriptive metadata is not to recreate the full metadata of an item we supply in the Digital Collections, but to provide baseline provenance of the item that can lead a patron back to our more thorough metadata sources. If someone finds a digital image shared outside of our Digital Collections, they can view the Content Credentials and easily determine:

1. that the item is hosted by Northwestern University Libraries
2. key metadata that contextualizes out-of-context items, and
3. where they can find more information about that item if needed

Key metadata fields for Content Credentials:

Host institution: for all items, this will be "Northwestern University Libraries"

- **Terms of Use**: all items will receive a terms of use statement, in line with the standard disclaimers used throughout the library in the dissemination of digitized material.
- **Title**: the title of the item, derived from the finding aid or catalog and pulled from the Meadow Title field.
- **Date**: the date the original item was created, pulled from the Meadow Date field.
- **Format**: the format of the original item, e.g. "black-and-white negatives" or "16mm film", pulled from the Meadow Genre field.

- **Creator**: the primary creator(s) of the original work, pulled from the Meadow Contributors field.
- **Item location**: the Alma number for catalog items or the collection name for special collections items, pulled from the Meadow fields Catalog Key or Related URL- Finding Aid.
- **ARK**: unique identifier automatically created by Meadow, which resolves directly to the item in Digital Collections when using a resolver tool. Crucially, this is a persistent identifier that will outlive Meadow in the case of future migrations.

Signing digital files (Kayla Warner)

NUIT has a certificate agreement through InCommon, but only for S/MIME certificates, not document signing certificates. Matt Stork has rights to generate S/MIME certificates now, but no one in NUIT or across the campus has done this before, so we (Matt) will need to figure out how to do this once we begin testing.

If we do need document signing, it will be more expensive and will require us to purchase it from a certificate provider. C2PA lists S/MIME as an acceptable type of signing. InCommon is free for us through NU, so using S/MIME through InCommon is probably the best first option.

Matt will be our best expert in figuring this out, and even he hasn't taken this on, but understands the language the best.

Next Steps

The next step will be to focus on the Pre-ingest work:

Create the Initial Pre-Ingest Manifest Store

How should metadata be compiled before a file is ingested into the repository? This will require substantial effort to standardize the data and map it to the C2PA schema. Ideally, each audio, image, or video file ingested into the repository would be associated with a C2PA-compliant manifest store. Coordinating with development and IT teams will be essential to support this work.

New Workflow Engine – Goobi

The Data Curation unit (which is comprised of digitization, metadata, and project management staff) currently uses SharePoint to track and manage its work. SharePoint has presented numerous challenges and offers limited support for automation.

Goobi is an open-source workflow management system designed specifically for libraries and archives, with current users including Stanford and Yale. In addition to tracking work, Goobi supports automated tasks and complex workflows. Its developers have worked closely with the community to build APIs for systems such as Alma and ArchivesSpace.

We would use Goobi to aggregate manifests and manage the Content Credentials workflow, with the goal of minimizing staff time spent on Content Credentials–related data entry and maintenance.

Future Actions

Once the pre-ingest steps have been established, we will be able to focus on post-ingest steps: How is the final version of the Content Credentials created and presented/delivered to the user? This will involve technical and design considerations.

Some steps in this process will involve:

- Descriptive metadata and publishing step of the workflow.
- Creation of a testbed
- Testing of signing functionality
- Further mock-ups of Content Credentials
- User testing and feedback
- Testing of editing of Content Credentials (the manifest store)

We will continue to engage with the community including:

Content Authenticity Initiative

C2PA | Verifying Media Content Sources

Digital Object Authenticity Working Group

C2PA for G+LAM headquartered at the Library of Congress

https://digitalpreservation.gov/meetings/DSA2025/010303_murray-potter_DSA2025-C2PA-G_LAM.pdf

Content Credentials Workflow
Northwestern University Libraries

Collection Materials*

**Born digital is not included in this model but will need to considered for the future*

Pre-Ingest

Intake
Evaluation
Inventorying

**Metadata is generated during this step that can be used later*

Reformatting
Capture
Post-Processing
Quality Control
Ingest

Generation of technical metadata from devices and software. Some of this equipment and software is C2PA compliant

Manifests

Post-Ingest

Metadata
Descriptive
Admin
Publishing

**Data gathered at intake is used during this step. Additional manifest is created and add to the store*

Manifests

File Processing
Transcoding
Streaming
IIIF Server

**The "A" access file goes through futher processing. The "P" preservation file is stored and will require its own manifest*

User Interaction

Library Website
Soft binding
Credentials are linked/referenced

Download
Hard binding
Credentials are inside the file

Some of the credentials information may be redundant on the Library website.

Mockups

There are 4 levels of visibility with C2PA. Level one is the “cr” logo indicating content credentials are present. Levels 2 and 3 reveal more information, and Level 4 contains the entire manifest store. Below are a few mockup with different display styles.