EDIT DRAFT

Content Authenticity Initiative for Heritage – A Primer

Table of Contents

Authorship and Scope

The coauthors of this primer are Julie McVey, Director of Digital Archives at the National Geographic Society (NGS) and Doug Peterson, Director of Research & Development at Digital Transitions (DT). Significant contributions have also been made by Ottar A.B. Anderson, Head of Photography at IKA More og Romsdal (IKAMR, the Norwegian Archive of Culture). This collaborative effort brings together industry knowledge of heritage digitization hardware and software with that of practitioners in the field representing digital archives and imaging services.

Because the goal of this primer is to acquaint heritage professionals and industry technologists with one another's work, it necessarily glosses over much of the nuance of imaging science and the philosophical debate over things like community trust and institutional authority. While these conversations are vitally important and expected to influence the direction content authenticity takes in the heritage community, for brevity they are outside the scope of this document.

Introduction

This document is intended to provide a primer on potential uses for content authenticity tools in the heritage digitization field, and to familiarize heritage professionals with the Content Authenticity Initiative (CAI) and its goals. The CAI is a large community of practitioners, from individuals to institutions, who have an interest in addressing disinformation across myriad industries. Members are not only technologists, but journalists, artistic creators, human rights activists–anyone with a vested interest in authenticating digital works, including cultural heritage professionals, can join this group. The CAI is focused on accelerating implementation of the C2PA open technical standard around digital content provenance. The Coalition for Content Provenance and Authenticity (C2PA) was launched by several technology companies in 2021 to work toward establishing an open technical standard specification for provenance. The C2PA sits within the Linux Foundation as a Joint Development Foundation Project, meaning its work is unencumbered by IP and is available for anyone and everyone to use. The CAI has developed a number of open-source tools based on the C2PA standard that may be incorporated into workflows for many different industries and fields, including photography, to establish authenticity of digital assets by writing and displaying content credentials.

By briefly explaining the motivations and methods of heritage digitization we hope to demonstrate how those could be served by implementing CAI (which refers to the process of implementing and displaying the C2PA's Content Credentials as well as the community itself). This primer advocates that technologists consider the needs of heritage professionals in designing and refining tools for content authenticity, and that heritage professionals actively consider how to incorporate CAI tools into their workflows.

What is Heritage Digitization?

The heritage community comprises primarily galleries, libraries, archives, and museums (GLAM) whose mandate is to collect, preserve, and disseminate our shared heritage. For

every object featured in exhibition spaces, these institutions hold thousands to millions more pieces of art, letters, photographs, natural history specimens, and countless other types of objects that can provide invaluable information to people worldwide.

In recent decades, digitization (or scanning[1]) of their physical collections has emerged as a core function of these institutions. While digitization has many goals, the most common is to provide a digital surrogate to replace the in-person use of the physical object, which both broadens access and serves preservation by limiting the wear and tear from physical use. Instead of traveling to the Library of Congress to read the first handwritten draft of Lincoln's Gettysburg Address, or to the Museum of Modern Art to view *The Starry Night*, a researcher can access a TIFF or JPG of these items on their respective websites. Digitization has had profound impacts on the way researchers and members of the general public can find and interact with the world's collections of heritage materials. As many institutions increasingly rely on digital surrogates to disseminate their collections and provide increased access to cultural heritage materials, concerns around the threats posed by disinformation and digital inauthenticity are growing.

With the explosion of information and disinformation in the modern cultural landscape, trustworthiness is an increasingly important core value for the heritage community. As the field strengthens its commitment to institutional integrity and earning community trust, institutions strive to establish and maintain credibility on many fronts. CAI tools could provide a valuable intervention with regards to trust of a digital object. To trust that an image of an object is a faithful and authentic representation of the object requires that the viewer know it has:

1. **Provenance Authenticity.** The image comes from a source the user trusts

[1] The term "digitization" is preferred. "Scanning" means using flatbed or planetary scanners that move ("scan") across the object. Modern digitization is typically camera-based. A more comprehensive primer on heritage digitization can be found here; anyone in the CAI community can contact dep@digitaltransitions.com for a code to take this class for free.

2. **Content Authenticity.** The image has not been modified between that source and the user (or any modifications are clearly described)

3. **Technical Authenticity.** The image has accurate color, tone, and detail. This is typically measured using an image quality guidelines or standards such as Federal Agency Digital Guidelines Initiative (FADGI) or ISO 19264

While institutions and professionals within the heritage community have long held these three tenets of authenticity as core values, there has not been an agreed-upon standard across the field for capturing and demonstrating them. We think that CAI tools could offer a standardized, traceable way to certify these key features of high-quality cultural heritage digital surrogates.

Content and Provenance Authenticity in Heritage

Content authenticity and provenance authenticity have been an issue in the heritage community far before the advent of digital imaging. For example, paintings can be forged, photographic film can be physically retouched, pages from a diary can be removed, or the authorship or provenance of an object can be misunderstood or intentionally misrepresented.

However, a digital representation of an object is far more easily modified (intentionally or unintentionally) than the physical original, and is outside the control of the source institution. Therefore, content authenticity for digital surrogates is a more extensive issue than it is for physical collections objects in heritage institutions. It is important to note that, much like authenticating physical objects, content authenticity for digital images does not prevent fraudulent reproductions, but instead provides a way to certify the authentic file as genuine.

There are at least four areas of concern heritage institutions have concerning content authenticity for digitized collections:

1. **Novel Forgeries:** The use of AI or traditional image manipulation to create a fabricated digital image that purports to be the authentic digitized image of a physical object.

2. **Inauthentic Image Modification:** The use of AI or traditional image manipulation to make changes to an otherwise authentic image.

3. **Metadata Modification:** The editing, removal, or addition of metadata in the image file, such as a change to the creator field.

4. **Interpretative Manipulation Tracking:** The presentation of a physical object sometimes requires practical image manipulation that should be documented and disclosed, such as the inversion of a photographic negative or contrast restoration of a faded document.

Technical Authenticity

Producing a high-quality digital reproduction of an object requires high-quality equipment that has been carefully calibrated. While many institutions used the best equipment available to them when starting their digitization programs, there were no agreed-upon standards for the precision and accuracy of color, brightness, detail, etc. For example, a simple Google Image search for *The Starry Night* shows myriad versions that differ in brightness, contrast, vibrancy, or detail. A more academic investigation of this variation can be found here; that paper represents the nascent era of digitization where best practices were not yet established and technical standards and methods to evaluate them were not yet developed.

To address the issue of technical image quality, in the 2010s, the heritage community began developing national guidelines (FADGI) and international standards (ISO 19264) that lay out metrics for image quality; physical targets that allow for the calculation of those metrics; and quality tiers for images that reach specific values for those metrics. For example, to achieve the highest FADGI rating, a rare book must be digitized at 400ppi or higher, retain 90% of the subject detail (SFR10 > 0.9) and must have extremely accurate

color (90th percentile of DeltaE2000 ≤ 4.0). A high-level overview of the workflow for measuring and verifying these metrics using FADGI is as follows:

1. A capture of a large Device Level Target is made
2. Software analyzes that capture and assigns the target image a quality rating from 1 to 4 stars[2]
3. Where it is practical, a smaller Object Level Target is placed alongside the material during image capture so each specific image can also be individually analyzed

These star ratings define whether the image is of sufficient technical quality for a given use case. For example, a 1-star image should only be used as a reference to locate the object in the collection, where a 4-star image can be used as a surrogate for nearly any in-person use. Institutions often use these star ratings as a measure of quality both in their own digitization and when outsourcing digitization to an outside vendor – the contract will require all delivered images conform to a specific star rating (usually 3-star or 4-star) and include the targets so that the institution can independently validate that quality level. Though technical standards and quality ratings have been widely adopted for heritage imaging, there is still no commonly accepted method for storing this metadata for longevity or communicating it to end users.

Source of Authority in Heritage Digitization

CAI is a way to confidently trace back an assertion of authenticity to some source of authority, and can be implemented at different points in the photography workflow. However, the best origin point of that assertion is dependent on context. For example, in photojournalism a photograph may be asserted as authentic at the time of capture by having the camera hardware itself sign the raw data of the sensor; the camera hardware is the source of authority. This is because the image is most "authentic" at that moment – the scene as recorded by the lens and sensor prior to any intervention or interpretation.

[2] There are several software and web-based services to analyze imaging targets, including OpenDice, Golden Thread, and Image Zebra.

In heritage digitization, the camera hardware is just one component in a system and workflow designed to produce an image that is an authentic surrogate for the original object. Additional non-camera-hardware components of that workflow include:

- **An "even field" capture** used to remove any vignette caused by the lens or unevenness in the illumination
- **A Custom Color Profile** that has been carefully calibrated based on the sensor, lens, illumination, and environment
- **Spectral Documentation** of the illuminant, the sensor, and in some cases of spots on the object
- **Image Quality Analysis** that has been performed on a target in the frame or separately captured during a preflight setup step
- **The Descriptive Metadata** that provides context to the image and ties it to the physical object

The quality standards and workflows required to achieve them add a second and equally important definition of "authenticity" to heritage imaging work. Therefore it is arguable that the most important stage to assert authenticity is not at the point of capture, but after the post processing that includes the above. These standard workflow steps ensure the most accurate representation of the physical object possible, and metadata about their inclusion in the process of creating the image provides an important bridge between the camera capture and the image as presented.

As CAI tools become more commonplace and integrated into photography workflows, organizations and individuals will need to decide the best place to begin the chain of authentication. If the camera hardware supports CAI, the raw sensor data can be signed by the camera hardware, and CAI allows that signature to be included in the final signed TIFF. But in heritage it is the heritage institution, not the camera manufacturer, that has the authority to assert that the image is an authentic representation of the original object.

Regardless of the raw authentication, the preservation-level TIFF must be asserted as authentic.

Finally, when the TIFF is subsequently used to generate smaller deliverable derivatives they can also be signed. For example the institution's Content Management System or Collections Management System (or other IT infrastructure that supports automation) can read and verify the CAI claims of the TIFF, and then create a JPG that lists that TIFF as an "ingredient" in the production of the JPG. In this way the end user of the JPG sees that the institution vouches for the authenticity of the JPG, and can also see that it was created from a TIFF that the institution also vouches for. This embedding of each prior asset's CAI claims as ingredients of the derivative provides traceability and transparency of the manner in which a particular asset was created. To see a variety of content credential examples, see C2PA's publicly available test file repository.

How CAI Can Help

The advent of cryptographic assertions of authenticity, tied to a chain of trust (certificates) provides an avenue by which institutions can positively assert the authenticity of their digitized objects.

CAI assures an institution that:

- Once the institution signs an image, any change to the image content or metadata will be shown as occurring *after* the signature and therefore not vouched for by the institution, unless re-signed
- No one can falsely sign an image as the institution

This does not prevent the act of image manipulation or metadata alteration (intentional or unintentional) itself. It is essentially impossible to prevent digital data from being modified – just ask the 1990s music industry. But any such change can be tied to a specific signature, so changes that are not tied to a signature are known to have happened after the most recent signature. Likewise CAI does not preclude a third party from using AI or image

manipulation from creating a fake image from scratch that they claim comes from an institution – it simply prevents that third party from signing it as if they were the institution. In short, CAI does not seek to detect fake images; it seeks to stamp authentic images. No one can prevent bad actors from making inauthentic images and saying they are from a given institution, but an institution that adopts CAI/C2PA can say that only images stamped by them as authentic should be trusted.

Because you can add metadata or ancillary data payloads to an asset prior to signing it, CAI also offers a platform for wrapping metadata, data files, or other image content that is related to the creation of the underlying asset. For example, the results from image quality analysis, spectral values of the camera and lighting, or an even field capture can be embedded and included in the information that is being cryptographically locked. Whether this is the best way to package that data requires additional scrutiny, but could prove a compelling use case for heritage photography.

While CAI holds promise as a technology that could be relatively easily incorporated into imaging workflows, there are obviously myriad questions the heritage community need to address surrounding the field's own technical standards and practices. There is also the fundamental question of trustworthiness, not only of the tech companies sponsoring these solutions and standards, but of cultural heritage institutions implementing them. Lastly, heritage institutions will also want to consider long-term digital preservation implications when adding anything to digital files.

Summary

- Heritage Community = Galleries, Libraries, Archives, and Museums (GLAM)
- Heritage users care deeply about authenticity
- Some heritage needs overlap those of other authenticity stakeholders

- Some heritage needs are specific to heritage – image quality results, target captures, even fielding, *input* color profiles

- CAI allows an institution to extend its authority to claims about a file (via a larger public chain originating with a Certificate Authority) The technical underpinnings of CAI are the same as online banking – extremely secure

- CAI is most important to apply at the TIFF and derivative stages; it can also be applied at the RAW stage if the camera supports it

- CAI may be an avenue to embedding metadata such as image quality analysis results

- CAI may be an avenue to embedding ancillary image content such as a target capture

Appendix: How CAI Technology Actually Works

The math and science behind CAI tools is somewhat involved. It is not necessary for every individual in an institution to understand the technical details for the institution to use CAI tools, any more than someone driving a car has to understand the details of how combusting gasoline drives the pistons. It's more important to understand the resulting benefit – using CAI an end user of an image can know:

- A file originates from the institution that signed it
- Whether an image has been altered after being signed

CAI does not seek to prevent a bad actor from making a fake or manipulated image of a museum/library/archive object. A bad actor could even use CAI technology to sign their fake image. But they cannot sign as the institution that holds that object. And if a bad actor manipulates an image signed by an institution, those modifications will be shown as taking place after it was signed.

A basic analogy can be helpful in understanding the underlying concept. Say John wants to be able to send Jane a note that she knows was not changed after he wrote it. John then makes an invisible ink that turns a very specific pink when lit by a very specific wavelength of UV flashlight and gives that flashlight to Jane. He then writes a note on plain paper with a plain pen: "Dear Jane, meet me at noon on Friday." He then writes on the note using invisible ink: "This note has four As, four Es, one I, three Os and no Us – John" and sends it to Jane. When she receives the note, even if it passes through the hands of other people, Jane simply uses the flashlight to see if a count of the vowels is accurate. If so, the note is from John and has not been changed. If the note has been changed she will know that it has been changed after John signed it, and can therefore be dubious of it.

All analogies are incomplete – the real-world math used for CAI means that:

- There are an *astonishing* number of [UV Flashlight + Ink] pairs; each ink can only be read by its paired flashlight.
- There is a third-party company that keeps a public record of which flashlight is associated with which person – but NOT the formulation of how to make the ink for it. When John gives Jane the flashlight Jane knows it wasn't a bad guy masquerading as John.
- Instead of a simple count-of-vowels the entire content of the letter is used to create the "hash" that is used to detect changes to the content.

For those interested in the technical terminology:

- The invisible ink is a **Private Key**
- The flashlight is a **Public Key**
- The count-of-vowels is a **Cryptographic Hash** such as MD5
- The third-party company that keeps track of flashlight ownership is a **Certificate Authority**

In short, it's essentially impossible to make an undetected change or to fake a signature. This is the same technology used for high-security banking – it's hard enough to defeat that anyone with the time/resources to try would surely be more interested in stealing from banks than forging the authenticity of a heritage image. Notably, the authority behind the claim of authenticity is ultimately the institution. So in the unimaginable case that someone did compromise the system (e.g. by stealing their ink / private key) the Director of the affected institution could simply make a press release saying they've been compromised and everyone will know not to trust files signed thereafter with their ink / private key. It is the institution that is the source of authority – CAI is simply a way to stamp files with that authority (and for end users to check that stamp) in an automated manner.

Verifying the authenticity of a file via CAI does not require the internet or a central authority to authenticate (though that is an option if desired); a file can be validated as authentic locally.