

MEDIA PROVENANCE – SIGNING YOUR CONTENT IN PRACTICE

D. J. Bevan¹, N. C. Earnshaw¹, L. C. Ellis¹, C. H. M. Halford¹, M. Hjelhaug²,
M. H. Iversen², M. F. Marcus¹, J. S. Parnall¹, L. Strapelli¹, H. O. Svella²

¹BBC, UK, ²Media Cluster Norway, Norway

ABSTRACT

Disinformation is a threat to the healthy functioning of democratic society. The advent of Generative AI technology means it is even easier for anyone to create false or misleading information and distribute it at scale making it almost impossible for people to trust anything they see online. This is a problem for most news organisations whose audiences trust their output less and less each year. How can organisations rebuild a trusted relationship with their audiences and consumers empowered to make their own decisions about the accuracy and veracity of what they see online?

Content credentials may be one solution. This technical standard exposes the origin and provenance of media, showing users exactly who published a piece of content and how it was made. This paper will summarise findings from public trials that assess the impact of content credentials on audience trust in news media. It will also review the adoption efforts at the BBC and in the Norwegian media industry.

INTRODUCTION

Trust in news has been declining for some time ‘Newman (1)’. There are many intersecting reasons for this, but recent advances in generative AI, and the fact that almost anyone can make synthetic (fake) images quickly and easily, are making it a lot more difficult for people to trust what they see online.

Studies have shown that it is currently very difficult for organisations and individuals to consistently detect what has been fabricated or manipulated to the very high accuracy level required for broadcast or publication ‘Chuangchuang et al (2)’. We therefore argue that it is better to look at a hybrid strategy that includes methods to positively assert who created the content and how it was made, and whether an AI tool was used. If you make this information available to the public, they can make an informed decision themselves whether to trust it. This strategy can include a combination of media provenance, watermarking and finger printing technologies.

As well as assertions from the original creator, content provenance is useful in cases where complex stories include contributions from multiple sources by making a 'chain of provenance' visible to the audience. This also has the benefit of increasing the transparency of the editorial process.

To ensure an interoperable and core set of signed media provenance features are available for all, the Coalition for Content Provenance and Authenticity (C2PA) has published an open standard describing a way to cryptographically sign content and associated standardised metadata 'C2PA 2024 (3)'. After signature verification, the metadata can be decoded by a "validator", such as a web browser extension, and displayed to viewers at a level of detail that they wish to receive. This capability is being rolled out under the user facing brand of Content Credentials, with adoption by a number of AI tools and early implementations by broadcasters.

There are many organisations actively investigating the value that C2PA provenance data can bring to their audiences and journalists and trialling how this new dimension can fit within their workflows. This paper covers trials for the practical use of content credentials from the BBC and Project Reynir – a collaboration led by Media Cluster Norway.

STANDARDS

There has been an awareness of the misrepresentation of published media for some time as the tools for manipulating media objects have become commonplace, just as the danger of images being separated from their original context has grown. In response, Project Origin was formed in 2019 by BBC Research & Development, CBC / Radio-Canada, Microsoft and the New York Times to consider a technological solution for application to the news media industry 'Aythya et al (4)'. In 2020 this group came together with the members of the Content Authenticity Initiative (CAI), led by Adobe, and together with other partners formed the Coalition for Content Provenance and Authenticity (C2PA) to develop an open standard for content provenance.

C2PA swiftly put together a plan for an initial "version 1" of the specification 'Earnshaw et al' (5), focussing on a few key areas:

- The assertions (or metadata) that the signers of content might want to assert about content (e.g. where a photo was taken, or its original caption)
- The provenance chain model (called "ingredients" in C2PA), allowing multiple stages of editing to be linked into a complete history
- The trust model, linking the assertions being made to the identity of the entity signing them.

Whilst the application of the Content Credentials can span a number of media-related verticals and industrial applications, Project Origin, now expanded in membership, remains focused on their application to news media and how to sustain robust and effective news distribution in the internet age using this technology.

How it works

The C2PA specification is fundamentally a way to do a few things:

1. Make (or "assert") some statements about the content, such as capture date, or description

2. Add links back to previous versions of the media, or its components, to optionally show the provenance of the media all the way back to the capture device
3. Add a cryptographic hash to the media, so that the provenance is tied to the media in a tamper-evident way
4. Add a tamper-evident digital signature (by the “signer”) over all the previous data, showing who or what made the provenance statements

This data is recorded in a structure called a Manifest, and multiple Manifests are included and linked, via 2, to show a full provenance history if required, in a “Manifest Store”. A Manifest Store can be embedded in a piece of media’s extension / metadata section (multiple “embeddings” are specified for different file types), or it can be stored in an internet-accessible location and a link to that location included in the media.

Finally, when some media is being consumed, all this data is accessed and read, and a “validation algorithm” is run to ensure the certificate that signed the media is valid, and all the links and the hashes are valid too. A “valid” Manifest can then be shown to the consumer, allowing them to see a secure description of what a “signer” says about the provenance of the content.

USER EXPERIENCE RESEARCH AND DESIGN

The BBC have been researching the user experience aspects of provenance since 2021. Our studies have helped us establish a greater understanding of:

- What image provenance information is important to show
- How people want to see that provenance information
- The impact on trust in content when provenance is shown

Some of this work has fed into the BBC’s live trial (below) where provenance information was disclosed in an expandable User Interface (UI) beneath image and video content on various BBC Verify articles. The following sections explain our research and subsequently our design decisions.

What information is important and relevant

We conducted a survey of 200 people as well as one-to-one interviews with 15 people to uncover qualitative and quantitative findings. We found that there are certain types of information that people are more interested in seeing when evaluating images, including: description, time, date, location, if the image has been verified, the verification checks conducted, who published the image, who the image is owned by, the creator/photographer, any image edits, and whether AI was used.

76% of respondents preferred a medium amount of information; 52% wanted additional information about the image; 89% agree/agree strongly showing the information is useful. It found that under 35’s were more concerned with contextual info e.g. creator/photographer, who published it, and who it was owned by, whilst over 35’s were more concerned with factual information: date time etc.

Another important aspect we found was the use of language when surfacing this information; some words have certain associations or can be perceived more negatively or positively. According to our qualitative research findings, there were a few key areas that require more exploration around the use of language: edits, verification and the use of AI.

- **Edits:** Participants from further interviews primarily wanted to know if an image had or had not been edited. Some wanted to know more about the kind of edits or be told if significant edits had been done, with the primary concern being misrepresentation or change of context, meaning or perception
- **Verification:** People primarily wanted to know if the image had been verified or not, and they were also interested in the kind of checks that were done to verify it. Ideally, they wanted these checks to be conducted by an independent body and wanted the ability to learn more about the verification process if desired.
- **Generative AI:** They primarily wanted to know whether or not generative AI was used in the creation of a particular image. People's level of digital literacy and use of language was a key factor in users' perceptions of this. Usage of generative AI was not viewed as acceptable in a news context, other than reporting on the use of AI. They also wanted this information to be flagged clearly to them.

How to show provenance information

We have been researching what design patterns people find most usable, helpful and engaging. C2PA recommend four levels of progressive disclosure, each with more detail:

- the first indicating the information is available
- the second with a summary of the key information users want
- and additional levels with considerably more detail.

A key challenge with disclosing provenance information is that there are large amounts of data which if shown all at once would be overwhelming and time consuming to have to read. People also have diverse needs; some are less interested in the details or unfamiliar with content production processes, while others are interested in interrogating the content further. Our research has focused primarily on the first and second levels, where key information is summarised. We have explored how to show information in a way that is objective but concise, while responding to the various user needs.

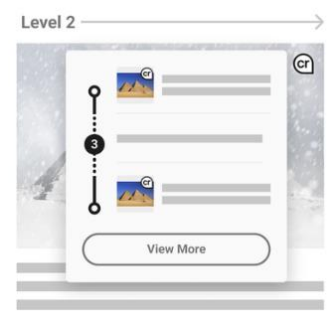


Figure 1 - level 2 display

We also explored potential layouts for presenting users with image provenance information. People wanted the information to be consistently displayed across devices and platforms and to be able to easily scan it to find key points of interest. However, they did acknowledge that this may not be possible given how different organisations may implement the technology on their platforms. From the research, we were able to pull out key visual principles that are important for people when seeing provenance information for images.

Medium amount of information – show all that is essential but more can be available.	Stay on the same page	Should be consistent across device/platform
See the image and text at the same time	Unobstructed image	Clear visual indication if the image used AI / was edited
Preference for drop down but expectation of pop-up	Link to further info.	Wanted bullet points

Table 1: Primary considerations for Visual Preferences

It's important to consider how presentation can impact people's perceptions and ensure that it is done in a way that does not influence them but allows them to make their own decisions. For example, people may want a system to flag content that may be misleading, using a traffic light system, however we know this can lead to misunderstanding and misrepresentation of the facts.

In the BBC trial below provenance information related to BBC Verify content and therefore we had to consider technical restrictions around what was possible to show to people, such as the available provenance data and web/app platform constraints. Our goal was to ensure the overall article experience was not negatively affected while making the provenance information clearly discoverable to people. We explored a number of possibilities within the constraints we had, and through testing these solutions we identified an approach using dropdown lists as the most effective and desired format.

How provenance information impacts trust

We found that surfacing this information through the methods and examples outlined above had a primarily neutral impact on trust, with some indication of instances of increased trust. This is true of users that do and do not use the BBC. Adding provenance information was also shown to increase trust in content published on BBC News for non-BBC users.

We conducted our tests based on three different types of images: editorial, stock image, and user generated content (UGC). Adding provenance information served as an equaliser of trust across the image types, with them all eliciting the same levels of trust. Where no provenance information is shown, UGC had the highest levels, followed by stock image and then editorial images.

There are still knowledge gaps around this space that we are looking to research further. These include education, the impact of the lack of consistency across sources, accessibility considerations, how to flag the use of AI and the impact on behaviour. Drawing on these learnings, we have built a lightweight way for journalists to sign and share the content credentials of key imagery we publish, giving audiences an insight into not just *what* we know, but *how* we know it. We have launched a live trial on the BBC News website to get a better and more representative understanding of whether this helps audiences increase their trust in news.

BBC TRIAL

Content credentials can help audiences discover how a piece of content was made and who made it. They help audiences assess whether a piece of content was made by the organisation who claims that they published it, and if they trust that organisation, they can trust the content. As more organisations use them, a larger proportion of authentic media will have content credentials included, and internet users will increasingly be able to judge the authenticity of what they are seeing or hearing for themselves.

It will take time for credentials to be widespread and visible across third party platforms, so we now describe a trial carried out in collaboration with BBC News. For the trial we looked at where we could use content credentials on our own platforms to add value to our content, by providing reassurance, transparency and more information about how we know what we know.

At the BBC, we strive to report the facts accurately, holding ourselves to the highest journalistic standards. BBC journalists do rigorous manual verification of the media we publish to ensure it is an authentic depiction of events. We do so by checking the content against other sources, examining the metadata, comparing locations, weather, and searching for other instances of the material online. We do everything we can to ensure we are not furthering the spread of disinformation. And where we do find fakes, we call them out. All this takes time, but we would rather be right than be the first to a story. We now have a dedicated team within the newsroom, BBC Verify, using a range of forensic investigative skills and open-source intelligence (OSINT) capabilities to validate incoming material.

“At BBC News we know that trust is earned. When our audiences know not just what we know, but how we know it, they feel they can trust our journalism even more.” – Deborah Turness, CEO of BBC News

We also recognise that it is not enough to ask audiences to trust us at face value and that we need to show audiences *how* we sort fact from fiction and do so securely, binding the *context* to the *content* so any tampering will be evident. This is why we have combined the cryptographic signing mechanism of C2PA with the details of manual verification from our journalists for images that come into the newsroom from sources lacking C2PA-enabled devices.

This not only means that audiences can make up their own minds about whether a piece of content is trustworthy, based on how much they trust the BBC, but also means that the provenance can be traced back to the BBC if the material is shared elsewhere.

“In a world of deep fakes, disinformation and distortion, this transparency is more important than ever.” — Deborah Turness, CEO of BBC News

Approach and Methodology

For the trial we worked with BBC Verify to integrate their work, carefully compiling their in-depth verification of the authenticity of an image or video into a claim that could be associated with the published media.

We focused on user generated content (UGC) because, from the research above commissioned by the BBC, it is the media people seem to trust the most, ahead of professionally captured media. It is also the most likely to arrive from an unverified source, and so requires manual verification before we can publish it as a piece of news. This process creates a proxy for the rich provenance data that could be provided directly by cameras and other devices in the future, and so we sought to use it to demonstrate the value of transparently surfacing context through provenance data.


To deliver content credential for BBC published material we built a C2PA-signing plugin that extends the BBC's internal UGC management tool that the BBC Verify team use to collect, collate, prepare, review, and approve provenance and authenticity metadata for UGC assets. The trial system was designed to add almost no additional work for journalists, integrating with existing systems and providing interfaces to give them editorial control.

Our design drew on the research done with BBC audiences to create an element that can be embedded into an article, like any other piece of media, showing the provenance data. For the trial we gathered feedback from users who interact with it about whether it affects the trust they have in it.

The plugin maps the manual verification data to C2PA assertions. As we are focusing on provenance claims, rather than inventing our own metadata schema, we adopted an existing review schema from schema.org - <https://schema.org/ClaimReview>. While we do not use every field in that data type, we do map as much of our data as fits that definition as possible. C2PA is a flexible standard which permits this form of flexible extendibility as a core feature.

The certificate we use to identify ourselves is issued by a Certificate Authority that is trusted by both the Content Credentials Verify and the Origin Verify tool –

<https://truepic.com/certificate-authority/> (see Certificate section). Should we elect to sign media that already contains existing C2PA credentials – for example, inserted and signed by a Leica or Sony camera body – the signing process can append our assertions to the manifest to achieve the “chaining of trust” as was earlier discussed.



Strike on military vehicle in Tetkino, Kursk

content credentials
Issued by BBC on Mar 12, 2024

About this video
Footage from Ukraine-based Russian paramilitary unit Freedom of Russia Legion claimed to show strike on Russian Armoured Personnel Carrier in the village of Tetkino, Kursk region, Russia.

Posted on
Telegram

Created
Mar 12, 2024

Location
51.274577, 34.281507 [View map](#)

Edits
Superficial edits were made to this content to improve technical quality, in line with editorial guidelines.

Verification checks
Completed by BBC Verify

The layout of roads, buildings and trees is consistent with publicly-available satellite imagery at this location. Green and blue roofs also evident on satellite imagery.

Onscreen caption at 7 seconds reads "enemy armoured personnel vehicle" in Russian

Reverse image searches on Google and Yandex search engines of three keyframes each returned no results - suggesting video has not been cached and is therefore a recent upload.

Shadow placement suggests footage was filmed early morning.

Weather conditions match those reported for this location on morning of 12/03/2024

Figure 2 - Article with Content Credentials on bbc.co.uk/news

To surface this metadata to audiences, we built a Content Credentials component that journalists insert into articles on the BBC News website.

We surface both the media asset, and its attendant provenance and verification metadata. The component features a hyperlink that opens the asset in the Content Credentials website - a page that features a C2PA “decoder”. Once decoded, the embedded C2PA metadata found in the file is validated. This site also checks that the metadata and image shown have not been tampered with or altered, as the C2PA bundle is signed against a hash of the media at the point of publishing. Any changes to metadata or content will immediately invalidate the integrity of the file, both on the Content Credentials website and on our own platform, and this is visible to the audience.

PROJECT REYNIR: THE SANDBOX FOR IMPLEMENTING C2PA

Project Reynir aims to secure an 80% implementation of C2PA in Norwegian newsrooms by the end of 2026. This will mean that Norway could be the first country with wide-scale implementation of the technology.

A pressing and practical issue for any industry seeking to implement C2PA is the extent to which companies spend a lot of time and money reproducing a set of processes across each individual organisation. In mitigating this challenge, Project Reynir is taking a holistic and collaborative approach, aiming to unite the efforts of the media industry in Norway. The main objective for the project is to establish an ecosystem of collaboration and the sharing of knowledge and experiences across an industry with strong individual actors.

Value Creation

Project Reynir is creating value through:

- **Accessibility:** Making the technology accessible as a common good that accelerates innovation and provides an industrial boost for the media industry.
- **Democratisation:** Granting small actors like the 200 local newspapers in Norway access to the same technology and industry standards as the world’s largest newsrooms
- **Expertise:** Securing expert knowledge and upskilling in the industry through the sharing of knowledge, technology and information.
- **Collaboration:** Contributing to the creation and development of value chains from academia, research and the private sector to end-users.
- **Innovation testing:** Providing a large-scale sandbox for proof-of-concept across a larger media ecosystem.

Collaboration and Partnerships

These objectives are being addressed through collaborative efforts that include developing shared standards, sharing best practices and creating a common platform for collaboration.

Key partners include editorial companies such as Schibsted, TV 2 Norway and NRK; technology firms like Vizrt and Factiveverse; and academic collaborators such as the University of Agder and Teklab at the University of Bergen. Other partners include the Norwegian news agency NTB and the fact-checking organization Faktisk, to mention just a few.

The first media tech companies in Project Reynir who are in the pipeline to support C2PA are Vizrt, CuttingRoom, Mimir and Wolftech. Vizrt is a world-leading provider of tools for visual storytelling in live broadcasts. They are seeking to implement C2PA for the purposes of live video, as well as for their MAM-system. CuttingRoom provides a cloud-based editing suite as well as a recording tool for journalists. They will implement C2PA for use in video production and editing. Mimir is a cloud-based media asset management system that are going to implement C2PA as a part of their pipeline in order to serve the media companies that use the service. Wolftech provides a workflow tool for journalists and will provide a seamless C2PA pipeline for the users.

There is a risk that the different companies will encounter a varying degree of complexity and face different challenges when implementing the technology. Some will support the viewing of content credentials for end users. Some will sign the content themselves, some will sign on behalf of their media company customers. Other companies deliver products that are involved in the production and processing of media assets. Others still are looking into new types of media assets and how the technology could be integrated here. By being part of the Project Reynir ecosystem, companies are able to build on each other's work, sharing knowledge obtained from working with the technology and assisting each other in the process.

Since the pre-study of Project Reynir kicked-off in August 2023, we have successfully managed to create substantial interest in the media industry, as well as built an ecosystem consisting of a variety of stakeholders including vendors, news media and academic institutions. We have aided in facilitating the technological implementation through connecting technology partners with the current CA, Truepic, to enable them to be first movers on implementing the technology. Our academic partners have already started conducting research projects into C2PA-relevant questions, as well as the application of the technology in a Norwegian context in collaboration with the newsrooms. Furthermore, after gathering all stakeholders and interested parties to the Project Reynir summit, we successfully managed to formalize the collaboration workflow and established a project structure for the road ahead.

In considering the enablement of wide adoption of C2PA in national market, we think that it has been an advantage that a neutral party, namely Media Cluster Norway, has facilitated the collaboration. Through Project Reynir, the entire industry has come together to solve a huge challenge and agree on a common way forward to implement Content Credentials through the chain. This approach could very well be reproduced and employed in other areas, be they countries, regions or particular industries.

MANAGING CERTIFICATES

In order to adopt Content Credentials, organisations need to obtain a certificate to sign their content. The International Press Telecommunications Council, in conjunction with Project Origin has established a working group to create and manage a C2PA-compatible list of verified news publishers. These publishers will be recognised by the system through a corresponding identification program.

The group has created the "Origin Verified Publisher" logo to convey the fact that content has been signed by a certificate granted to a publisher that has been verified according to the Project Origin process.

Origin Verified Publisher Certificates will ensure that the identity of established news organisations are protected from imposters. The certificates confirm organisational identity and do not make any judgement on editorial position. Liaison agreements with other groups in the media ecosystem will be used to accelerate the distribution of certificates.



Figure 3 - Origin Verified Publisher Logo

The initial trust list uses Truepic as a certificate authority, with the BBC and CBC/Radio-Canada as trial participants. In the early phases of trust list roll-out, the IPTC Media Provenance Committee will be designing the policies required to issue a certificate, as well as slowly expanding the trial to include more organisations. These are likely to be drawn from IPTC members in the early trial phase, with the final intention being that it is possible for any organisation (regardless of IPTC membership status) to request a certificate.

CONCLUSION

Media Provenance is being trialled in BBC and Norway and the adoption is increasing within both news organisations, vendors and platforms, including creators work using of generative AI. Initial work focuses on the point of publication (signing media and metadata from a publisher), as there is not yet enough adoption for realistic “glass to glass” (e.g. signing media from the point of capture through to credential display on a user device) workflows. This is expected to change rapidly as the ecosystem evolves.

Early trials are showing success in the development of trust relationships and helping users understand which media to trust, and these will be explored in more detail as trials continue. However, challenges remain in helping users understand the signals from C2PA provenance and there will need to be a concerted effort to explain to the concepts to general users.

The Trust list, a critical part of the ecosystem, is at an early stage but is getting ready to expand to C2PA adopters in news.

Media Provenance has the potential to make significant impact on the challenge increasing trust in news and the security of other media online.

References

1. Newman, N. 2023. Digital News Report. Reuters Institute. 2023
2. Chuangchuang Tan, Huan Liu, Yao Zhao, Shikui Wei, Guanghua Gu, Ping Liu, Yunchao Wei, 2023. Rethinking the Up-Sampling Operations in CNN-based Generative Network for Generalizable Deepfake Detection. ArXiv. December 2023
3. C2PA Specifications, 2024. C2PA website, <https://c2pa.org/specifications/specifications/2.0/index.html> 2024
4. Aythora J., Burke R., Chamayou A., Clebsch S., Costa M., Earnshaw N., Ellis L., England P., Fournet C., Gaylor M., Halford C., Horvitz E., Jenks A., Kane K., Lavallee M., Lowenstein S., MacCormack B., Malvar H., O'Brien S., Parnall J., Shamis A., Sharma I., Stokes J., Wenker S., Zaman A.. 2020. Multi-stakeholder Media Provenance Management to Counter Synthetic Media Risks in News Publishing. International Broadcasting Convention. September 2020.

5. Fighting Misinformation with Authenticated C2PA Provenance Metadata, 2023. Earnshaw N., Dupras J., MacCormack B., NAB Broadcast Engineering and Information Technology Conference. April 2023

ACKNOWLEDGEMENTS

The authors would like to thank their colleagues for their contributions to the work supporting this paper. We would like to thank BBC Verify and BBC Product for enabling us to implement the trial on the BBC News website. We would also like to thank Truepic for the Certificate Authority, Microsoft for all their support in development, CBC / Radio-Canada, for joining us in launching the Origin Trust List and the Content Authenticity Initiative, for their open-source C2PA library.

They would also like to thank the International Broadcasting Convention for permission to publish this paper.