IP SQUARED

DIGITAAL WERKTUIG

Rapport Verkenning implementatie C2PA

NPO Innovatie, Media Campus NL en Beeld & Geluid, juni 2025

Colophon

Opdrachtgevers – NPO Innovatie, Media Campus NL en Beeld & Geluid

Auteurs – Maarten Zeinstra (IP Squared), Maarten Brinkerink (Digitaal Werktuig)

"Zeinstra, M. & Brinkerink M. (2025), C2PA n'est pas une Pipe (Rapport Verkenning implementatie C2PA), *Nederlandse Publieke Omroep."*

AI transparantie statement – Er is van generatieve AI gebruikgemaakt tijdens de ontwikkeling van dit document. Het document bevat alleen tekst die door een menselijke eindredactie is nagelopen en geverifieerd.

Afbeelding omslag – De afbeelding op de omslag is gegenereerd door Gemini Advanced 2.0 Flash, gebaseerd op de prompt *"I want to generate an image based on a parody. The Image is about the open C2PA standard. I want to make a parody that states something like "C2PA n'est pas une pipe" referencing 'Ceci n'est pas une pipe' By René Magritte"*. De afbeelding is nadien bewerkt, waaronder het vervangen van de tekst boven de afbeelding.

Figuur 1. Originele afbeelding uit Gemini. Deze had geen C2PA-manifest.

Management summary (English)

This document is written in Dutch, except for this management summary and the summaries of the interviews in the Annexes with Dr. Neal Krawetz, Shawn Masters, Lucille Verbaere and Mohamed Badr Taddist.

A Dutch management summary can be found below.

Background and Objective – Commissioned by NPO Innovatie (NPO), Media Campus NL, and Beeld & Geluid, this research investigated whether the C2PA (Coalition for Content Provenance and Authenticity) specification is suitable to support NPO's objective in this study: *"To combat disinformation by ensuring the authenticity of reporting through digital certificates."* This investigation builds upon the previous Proof of Provenance project (2022) and considers the rapid developments surrounding C2PA and the EBU's call to support standards for origin verification.

What C2PA Offers – C2PA provides a technical method to link *provenance* information via a digitally signed 'manifest' to media files (assets). This makes it possible to verify whether the file and its associated claims (e.g., about the creator, edits) have been altered since signing. It can be applied in various stages: production (by creators, editorial teams, AI tools), distribution (integrity check), and consumption (information for the public, GenAI detection).

Core Finding ("The Treachery of Images") – The report concludes, analogous to Magritte's caption "Ceci n'est pas une pipe," for his painting "The Treachery of Images", that C2PA is a *representation* of authenticity and provenance, but **offers no guarantee of the truth or objectivity of the content itself**, **nor of the factual accuracy of the provenance claims within the manifest**. It merely confirms that a specific signer stands behind the claims and that the file has remained unchanged since then. Trust in the media expression remains dependent on the consumer's trust in the signing party (the 'brand') and in the technology. C2PA is therefore a **valuable tool for transparency**, but **not a silver bullet against disinformation**.

Critical Notes and Risks – The report identifies several points of criticism and risks:

1. **Technology:** Vulnerabilities such as potential signature forgery, dependence on C2PA's own certificate authorities, insufficient validation of 'claims within claims', privacy risks associated with verification tools, and the (current) lack of a mandatory link to a verifiable digital identity.
2. **Governance:** The dominance of large tech companies in the C2PA coalition raises questions about interests and future direction. Potential conflicts of interest in the ISO standardization process.
3. **Work Processes:** Implementation requires careful integration into existing workflows to avoid friction and errors. Incorrect use can lead to reputational damage (see examples involving BBC, CBC).
4. **End User:** The added value of merely a 'checkmark' for the consumer is questionable; transparency about the *creation process* is seen as potentially more valuable.

Recommendations for a media publisher media publisher – If a media publisher decides to implement C2PA, the following steps are recommended:

1. **Strengthen the Specification:** Advocate for improvements such as mandatory digital identity and transparent certificate chains, possibly via the EBU context.
2. **Define Clear Objectives:** Define a limited, clear scope for the use of C2PA within the media publisher.
3. **Start Small:** Consider an initial implementation like automatically signing content on official media publisher distribution channels (e.g., own website, distributors), rather than the entire production chain immediately.
4. **Communicate Clearly:** Explain unambiguously to staff and the public what a C2PA signature means (and what it doesn't, such as not being a claim of truth).
5. **Introduce Own Verification:** Develop or use a publicly-controlled verification tool to mitigate privacy risks and maintain control over trusted sources ('trust list').
6. **Review Processes:** Analyze and document how C2PA fits into editorial and distribution processes.
7. **Increase Transparency (General):** In any case, provide more insight into how the media publisher/broadcasters verify sources, independent of C2PA.
8. **Consider Alternatives:** Also look into simpler standards like SEAL for specific purposes such as monitoring integrity during transport.

Conclusion – C2PA can be a useful instrument in the fight against disinformation by increasing transparency about provenance, provided its limitations are recognized, implementation is carried out carefully, and communication to the user is clear. It is a tool, not a complete solution.

Managementsamenvatting

Achtergrond en doelstelling – In opdracht van NPO Innovatie (NPO), Media Campus NL en Beeld & Geluid is onderzocht of de C2PA (Coalition for Content Provenance and Authenticity) specificatie geschikt is om de doelstelling van de NPO in dit onderzoek te ondersteunen: *"Desinformatie te bestrijden door de authenticiteit van berichtgeving te borgen middels digitale certificaten."* Dit onderzoek bouwt voort op het eerdere Proof of Provenance project (2022) en houdt rekening met de snelle ontwikkelingen rond C2PA en de oproep van de EBU om standaarden voor herkomstverificatie te ondersteunen.

Wat C2PA biedt – C2PA biedt een technische methode om *herkomstinformatie* (provenance) via een digitaal ondertekend 'manifest' aan mediabestanden (assets) te koppelen. Dit maakt het mogelijk om te verifiëren of het bestand en de bijbehorende claims (bijv. over maker, bewerkingen) sinds de ondertekening niet zijn gewijzigd. Het kan worden toegepast in verschillende fasen: productie (door makers, redacties, AI-tools), distributie (integriteitscheck) en consumptie (informatie voor publiek, GenAI-detectie).

Kernbevinding ("La Trahison des images") – Het rapport concludeert, analoog aan Magritte's het onderschrift "Ceci n'est pas une pipe", bij het schilderij "La Trahison des images", dat C2PA een *representatie* is van authenticiteit en herkomst, maar **geen garantie biedt voor de waarheid of objectiviteit van de inhoud zelf**, **noch voor de feitelijke juistheid van de herkomstclaims in het manifest**. Het bevestigt enkel dat een specifieke ondertekenaar achter de claims staat en dat het bestand sindsdien ongewijzigd is. Het vertrouwen in de uiting blijft afhankelijk van het vertrouwen van de consument in de ondertekenende partij (het 'merk') en in de technologie. C2PA is daarmee een **waardevol instrument voor transparantie**, maar **geen wondermiddel tegen desinformatie**.

Kritische noten en risico's – Het rapport identificeert diverse kritiekpunten en risico's:

1. **Technologie:** Kwetsbaarheden zoals mogelijke vervalsing van handtekeningen, afhankelijkheid van C2PA's eigen certificaatautoriteiten, onvoldoende validatie van 'claims binnen claims', privacyrisico's bij verificatietools, en het (momenteel) ontbreken van een verplichte koppeling aan een verifieerbare digitale identiteit.
2. **Governance:** Dominantie van grote techbedrijven in de C2PA-coalitie roept vragen op over belangen en toekomstige richting. Mogelijke belangenverstrengeling bij het ISO-standaardisatieproces.
3. **Werkprocessen:** Implementatie vereist zorgvuldige integratie in bestaande workflows om frictie en fouten te voorkomen. Onjuist gebruik kan leiden tot reputatieschade (zie voorbeelden BBC, CBC).
4. **Eindgebruiker:** De toegevoegde waarde van enkel een 'vinkje' voor de consument is twijfelachtig; transparantie over het *totstandkomingsproces* wordt als potentieel waardevoller gezien.

Aanbevelingen voor mediapartijen – Indien een mediapartij C2PA wil inzetten, worden de volgende stappen aanbevolen:

1. **Verstevig de specificatie:** Zet in op verbeteringen zoals verplichte digitale identiteit en transparante certificaatketens, eventueel via EBU-verband.
2. **Bepaal duidelijke doelstelling:** Definieer een beperkte, heldere scope voor de inzet van C2PA binnen de mediapartij.
3. **Begin klein:** Overweeg een initiële implementatie zoals het automatisch tekenen van content op officiële distributiekanalen (bijv. eigen websites of online distributies), in plaats van direct de hele productieketen.
4. **Communiceer helder:** Leg eenduidig uit aan medewerkers en publiek wat een C2PA-handtekening betekent (en wat niet, zoals waarheidsclaim).
5. **Introduceer eigen verificatie:** Ontwikkel of gebruik een eigen verificatietool om privacyrisico's te mitigeren en controle te houden over vertrouwde bronnen ('trust list').
6. **Licht processen door:** Analyseer en documenteer hoe C2PA past in de redactionele en distributieprocessen.
7. **Verhoog transparantie (algemeen):** Maak sowieso meer inzichtelijk hoe de de media partij/omroepen bronnen verifiëren, los van C2PA.
8. **Overweeg alternatieven:** Kijk ook naar eenvoudigere standaarden zoals SEAL voor specifieke doelen zoals integriteitsbewaking tijdens transport.

Conclusie – C2PA kan een nuttig instrument zijn in de strijd tegen desinformatie door transparantie over herkomst te vergroten, mits de beperkingen worden onderkend, de implementatie zorgvuldig gebeurt en de communicatie naar de gebruiker helder is. Het is een hulpmiddel, geen totaaloplossing.

Inhoud

Colophon 2
Management summary (English) 3
Managementsamenvatting 5
Inhoud 7
Introductie 8
Provenance en Authenticiteit 10
Provenance 10
Authenticiteit 10
Beoordeling herkomst en authenticiteit door de consument 12
Use cases C2PA mediasector 13
Productie 13
Distributie 16
Consumptie 16
Archivering 17
Kritische noten C2PA 19
Technologie 19
Coalitie en governance 21
Werkprocessen 23
Eindgebruiker 23
Mitigerende stappen 25
Bevindingen – La Trahison des images 27
Aanbevelingen 27
Bijlagen 29
Bijlage 1 – Samenvatting Proof of Provenance (PoP) 30
Bijlage 2 – Technische uitleg C2PA 32
Bijlage 3 – Interviews 34
Bijlage 4 – Secure Evidence Attribution Label (SEAL) 49
Bijlage 5 - Beschikbare tools en implementaties 50

Introductie

De Stichting Nederlandse Publieke Omroep (NPO) heeft in 2022 het Proof of Provenance project afgerond[1], in samenwerking met partners binnen het Public Spaces netwerk. Het project richtte zich op het vertrouwen in publieke media, en beoogde met digitale handtekeningen de authenticiteit van online nieuwsberichten te kunnen garanderen. In deze conceptfase 2022 is een werkende demo ontwikkeld. De NPO onderzoekt of en welke vervolgstappen gezet kunnen worden. Mede gebaseerd op de resultaten van het eerdere project, inclusief een SWOT and GAP analyse door IP Squared[2].

Sinds de afronding van het Proof of Provenance project is met name de adoptie en functionaliteit van de specificatie[3] van de Coalition for Content Provenance and Authenticity (C2PA) stevig ontwikkeld. In augustus 2024 heeft het Technical Committee van de European Broadcasting Union (EBU) een oproep aan de media-industrie gedaan om standaarden voor het verifiëren van de herkomst en authenticiteit van media-uitingen te ondersteunen, om zo hun publiek te helpen om de oorsprong van deze uitingen te achterhalen. Hierbij werd de C2PA specificatie bij naam en toenaam als voorbeeld van een dergelijke 'standaard' genoemd. NPO Innovatie, Media Campus NL en Beeld & Geluid willen de adoptie van C2PA nader onderzoeken. Hierbij vragen ze zich af of C2PA geschikt is voor het doel van de NPO met dergelijke technologie.

> *"Desinformatie te bestrijden door de authenticiteit van berichtgeving te borgen middels digitale certificaten."*

Ook in deze vervolgstappen beogen de partners desinformatie te bestrijden door de authenticiteit van berichtgeving te borgen middels digitale certificaten. Dit rapport onderzoekt of Authenticiteitcertificaten met de C2PA specificatie fit for purpose zijn met deze doelstelling.

Dit rapport begint met een theoretisch kader over herkomstinformatie en authenticiteit. Vervolgens wordt aan de hand van use cases beschreven hoe C2PA ingezet kan worden om in de levensloop van een media-uiting. Vervolgens worden er kritische noten bij de (inzet) van de specificatie geplaatst. Het rapport eindigt met een samenvatting van de bevindingen en een serie aanbevelingen voor volgende stappen.

[1] Zie Bijlage 1 – Samenvatting Proof of Provenance (PoP)
[2] Zie SWOT and GAP Analysis of the Proof of Provenance project – PublicSpaces.net
[3] Een specificatie lijkt op een standaard, maar zonder dat deze geaccepteerd is door een standaardautoriteit (zoals de ISO). De termen worden wel eens door elkaar gebruikt. Voor dit rapport gebruiken we specificatie. Zie Standards and specifications – what exactly are they? | U.S. Geological Survey Tenzij anders aangegeven wordt in dit rapport met C2PA verwezen naar de specificatie niet de coalitie.

De auteurs baseren zich op uitgebreid bureauonderzoek en een serie interviews met experts over de technische aspecten, de journalistiek en de mediasector. In het rapport wordt rechtstreeks naar deze interviews verwezen. Een verslag van elk interview is te vinden in bijlage 3[4]. De volgende mensen zijn geïnterviewd:

1. Tom Demeyer, Voormalig CTO Waag Society
2. Dr. Neal Krawetz and Shawn Masters, veiligheidsonderzoekers
3. Lucille Verbaere and Mohamed Badr Taddist, EBU C2PA Workflows
4. Prof. dr. Yael de Haan, Bijzonder hoogleraar Lokale Publieke Omroep
5. Prof. dr. ir. Jeroen de Ridder, Full Professor, Faculty of Humanities, Epistemology and Metaphysics

We bedanken deze experts voor hun tijd en vele inzichten.

[4] Zie Bijlage 3 – Interviews

Provenance en Authenticiteit

De Coalition for Content Provenance and Authenticity (C2PA) en de gelijknamige technische specificatie bieden een mechanisme om herkomstinformatie (provenance) verifieerbaar aan een media-expressie te koppelen. Producenten, beheerders, distributeurs en gebruikers van content kunnen daarmee verifiëren of het materiaal en de herkomstinformatie sinds de verspreiding niet aangepast zijn. Daarnaast kan deze herkomstinformatie beschrijvingen van bewerkingen op het mediabestand (asset) bevatten. De coalitie brengt de specificatie aan de man door het creëren van een ecosysteem van digitale herkomst-compatibele applicaties en tools.[5]

De mogelijkheid om de herkomst van media-uitingen te achterhalen en te verifiëren speelt in op maatschappelijke uitdagingen rondom vertrouwen in media, misinformatie, desinformatie en AI genereerde media-uitingen. Bij al deze uitdagingen spelen herkomst en authenticiteit een cruciale rol. Voor de daadwerkelijke analyse van C2PA is het daarom eerst van belang om duidelijk te krijgen waar het eigenlijk over gaat: Wat is provenance en wat is authenticiteit in het kader van C2PA?

Provenance

Provenance, ook wel herkomstinformatie, is informatie die aangeeft wie of wat de bron is van een bepaalde uiting of handeling. Dit kan zijn wie de makers zijn, wie de informatie verspreid heeft, wie het aangepast heeft, etc. Bij provenance gaat het om de beschrijvende informatie, of metadata.

In het kader van C2PA ziet dit deel van de specificatie toe op de informatie, de *claims* en de *assertions*.[6] Denk hierbij aan de technische standaarden en specificaties die voorschrijven hoe beschrijvende informatie in verschillende soorten mediabestanden gezet kan worden. Metadatastandaarden als EXIF, IPTC en XMP, maar ook bestandsformaatspecifieke informatie bij jpg, png, GIF, TIFF, worden genoemd. Ook externe unieke identifiers zoals URN worden aangehaald in de specificatie. Net als locatie-informatie en velden voor datum- en tijd en zelfs informatie over gebruikte lettertypes.[7] Veel van deze informatie kan door de editor zelf ingesteld worden. Zij kunnen er bijvoorbeeld voor kiezen om voor de (vrije tekst) waarde van het auteursveld een eigen naam te gebruiken, die van de organisaties, of zelfs een fictieve (en eventueel misleidende) naam.

Authenticiteit

Authenticiteit is een complexer begrip. Iets is authentiek als het trouw is aan zichzelf. In het kader van deze specificatie is authenticiteit een bevestiging van herkomstinformatie door de ondertekenaar.

[5] Zie Bijlage 5 - Beschikbare tools en implementaties
[6] Zie Bijlage 2 – Technische uitleg C2PA
[7] Zie Content Credentials : C2PA Technical Specification

De technologisch verifieerbare authenticiteit van C2PA heeft een meer rigide betekenis dan een merk. Om de authenticiteit van informatie technologisch verifieerbaar te maken zijn methoden nodig om te controleren of de informatie na de ondertekening niet meer gewijzigd is.

Authenticiteit en merken

Authenticiteit is bijvoorbeeld te beschrijven aan de hand van een voorbeeld van een tas die verhandeld wordt op de markt. Een authentieke tas is niet alleen trouw aan het ontwerp, maar is ook daadwerkelijk gemaakt door het merk dat het claimt te hebben. Een namaak-tas lijkt misschien op de echte, maar is niet door de eigenaar van het merk gemaakt. Het is daarmee niet authentiek.

Authenticiteit lijkt dan ook veel meer op de manier waarop wij merken gebruiken in de markt. Merken worden gebruikt om verwarring te voorkomen tussen soortgelijke producten. Zoals duidelijk onderscheid te kunnen maken tussen laptops van Apple en Razer of Dell. Een merk, zoals de NOS, wordt gebruikt om de herkomst van een product te duiden. De gebruiker kan dan zelf – op basis van (subjectieve) waardering van dit merk – bepalen of ze dit product voldoende vertrouwen en willen aanschaffen of consumeren.

Bij media-uitingen, met name de journalistiek, is deze authenticiteit complexer geworden. Prof. dr. de Haan wijst op de gelaagde identiteit van de afzender van een media-uiting. Authenticiteit is niet langer alleen maar gekoppeld aan het merk van de uiting, zoals de NOS, Argos, etc. Door social media is de journalist of presentator zelf ook een merk of teken van authenticiteit geworden. Met name Follow the Money en de Correspondent zetten de journalisten zelf als merk van authenticiteit in. De associatie van een mediaconsument met het "personal brand" van een individuele journalist kan anders zijn dan de associatie met zijn/haar werkgever (en zelfs botsen).

Net als bij een merk in de markt, wordt de reputatie van de partij die het C2PA-manifest ondertekend heeft, (al dan niet) vertrouwd. In tegenstelling tot merken, poogt C2PA ook technische zekerheid over de herkomst van de informatie te bieden. We weten wat er in een manifest staat, hoe deze opgebouwd is en hoe deze bevestigd is. De bevestiging (de digitale handtekening) geeft C2PA de autoriteit en vertrouwen in het manifest. Aangenomen vertrouwen in de technologie zorgt voor een vertrouwen in de verklaring van de specificatie.

Prof. Dr. Ir. de Ridder geeft aan dat dit een heel normaal proces is in de samenleving. Denk bijvoorbeeld aan onze omgang met bankbiljetten. Deze dragen technische middelen of kenmerken met zich mee om te bevestigen dat iets ‘echt’ geld is. Ook dan is er een autoriteit die aangeeft hoe geld eruit ziet. En er is vertrouwen in die autoriteit. De vraag is of C2PA als technisch middel hetzelfde vertrouwen als autoriteit gaat genieten.

C2PA wordt in het Engels ook weleens uitgelegd aan de hand van een analogie van het ‘notarizing’ (**notariseren**) van documenten. De vertaling naar het Nederlands veroorzaakt een anglicisme waar voorzichtig mee omgegaan moet worden!

‘To notarize’ betekent dat een persoon die openbare macht vertegenwoordigt kan bevestigen dat een handtekening op een document gezet is door de ondertekenaar. Dit moet niet verward

worden met de Nederlandse notaris, die additionele taken vervult (zoals het bewaken van belangen van betrokkenen en het beheren van authentieke akten die als onbetwistbare waarheid gelden). Hierdoor kan de verwarring in het Nederlands ontstaan dat het Engelse notarize ook over waarheid of objectiviteit kan gaan.

> "authenticiteitsmechanismes zijn geen shortcut naar de waarheid"[8]

Een C2PA-manifest bevat geen controle van de inhoud van het document en is daarmee ook geen controle op waarheid of objectiviteit, of de inhoud van herkomstinformatie zelf. Bij C2PA wordt deze taak door de ondertekenaar zelf uitgevoerd. Wil deze zichzelf (bewust of onbewust) als afzender van misinformatie of desinformatie bevestigen, dan is de ondertekenaar daar vrij toe.

De specificatie waarschuwt hier ook voor:

> *"C2PA specifications SHOULD NOT provide value judgments about whether a given set of provenance data is 'good' or 'bad,' merely whether the assertions included within can be validated as associated with the underlying asset, correctly formed, and free from tampering."*[9]

Beoordeling herkomst en authenticiteit door de consument

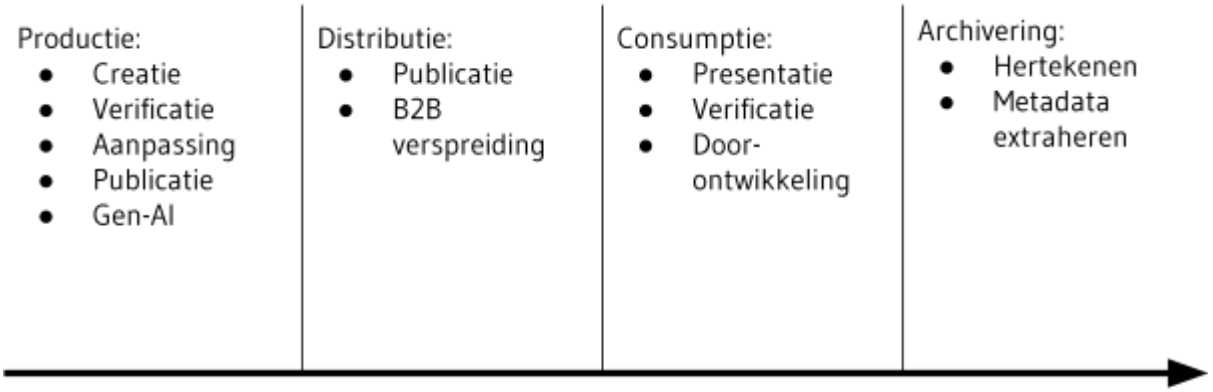
Het oordeel of de herkomstinformatie en authenticiteit vertrouwd kan worden ligt niet bij de technologie, maar bij de mediaconsument. Uit de interviews blijkt dat het kritische beoordelingsvermogen van deze consument niet onderschat moet worden. Te veel nadruk op het (vanuit een mediapartij) duiden van de 'betrouwbaarheid' van de informatie, bijvoorbeeld middels factchecking, kan zelfs een tegengesteld effect hebben, werd door geïnterviewden opgemerkt. Verspreiding en resulterende consumptie van misinformatie staat volgens verschillende geraadpleegde experts niet gelijk aan slecht of verkeerd geïnformeerde burgers.[10] Hierop aansluitend komt het wantrouwen van (sommige) mediaconsumenten volgens De Haan niet (enkel) voort uit ontbrekend bewijs van authenticiteit van media-uitingen, maar eerder uit een gebrek aan representativiteit, de complexiteit van "waarheid" en de veelheid van perspectieven in een 'post-truth' samenleving.[11]

Verschil mis- en desinformatie.
Bij Misinformatie wordt er feitelijk aantoonbaar onjuiste informatie verspreid. Bij desinformatie wordt dit aantoonbaar intentioneel gedaan. Desinformatie aantonen is lastiger dan misinformatie aantonen, omdat dit de (bewust deceptieve) intenties van de verspreider behelst.

[8] Zie Prof. dr. ir. Jeroen de Ridder
[9] Specificatie 1.2 'Scope' Content Credentials : C2PA Technical Specification
[10] Zie Prof. dr. Yael de Haan en Prof. dr. ir. Jeroen de Ridder
[11] Zie Prof. dr. Yael de Haan

Use cases C2PA mediasector

In dit hoofdstuk wordt C2PA toegepast in de verschillende stappen van de levensloop van media-uitingen (zie figuur 2). Dit is uitgewerkt in een aantal use cases die in de context van media beschrijven welke rol C2PA specifiek en "content authenticity" in het algemeen hierin kan spelen. Hierbij wordt ook ingegaan op de huidige uitdagingen bij het implementeren van deze use cases (waarover meer in het volgende hoofdstuk[12]).



Figuur 2. Levensloop media

Productie

Desinformatie bestrijden en de authenticiteit van berichtgeving borgen begint bij de productie van media-uitingen. Desinformatie en valse berichtgeving kunnen aan het begin van de levensloop ontstaan, door misinformatie of desinformatie over te nemen in eigen berichtgeving. Misinformatie of desinformatie kan ook ontstaan op het moment van bewerking door de redactie. Het is hierbij van cruciaal belang dat de menselijke processen rondom C2PA goed georganiseerd zijn en in de gehele keten nageleefd worden.

Dit in het achterhoofd houdende zijn er verschillende use-cases denkbaar op het moment van de productie van een media-uiting:

De maker tekent zijn werk – Een fotograaf, verslaggever, cameraman of geluidsman kan hardware gebruiken dat C2PA ondersteunt. Hierdoor wordt de opname direct ondertekend op het moment van opname.[13] Doordat bronmateriaal ondertekend wordt kan er vertrouwen geschapen worden wanneer, waar en hoe het ruwe materiaal van de uiteindelijke media-uiting tot stand is gekomen.

Dit komt met een risico: Een bijdrager aan de media-uiting die om moverende redenen anoniem wil blijven kan mogelijk toch identificeerbaar worden via het C2PA-manifest dat (automatisch) wordt gegenereerd tijdens de productie van de uiting, wanneer dit proces van C2PA

[12] Zie Kritische noten C2PA
[13] Zie C2PA Camera Support

ondersteuning voorzien is. Overigens kan dit ook via andere standaarden en technologieën gebeuren.

Naast hardware-implementaties bestaat er ook software voor het toevoegen van C2PA ondersteuning aan applicaties. TruePic biedt bijvoorbeeld APIs waarmee het C2PA protocol toegevoegd kan worden aan software op Android en iOS.[14] Uiteraard beschermt dit niet tegen bad actors die deze firmware of de gebruikte software zouden kunnen manipuleren. Een niet complete lijst met implementaties is te vinden in bijlage 5.[15]

Ten slotte is het mogelijk om mediaproducties die vervaardigd of bewerkt zijn met software die C2PA niet ondersteunt voor distributie 'handmatig' van een C2PA handtekening te voorzien. Dit vereist wel technische kennis voor het toepassen van deze (open source) software en begrip van de standaard en opbouw van een C2PA-manifest, om zo de kwaliteit van de ondertekening te waarborgen.

Maker zoekt naamsvermelding – Vanuit de maker geredeneerd is het primaire doel van het zetten van een dergelijke handtekening – in de context van de C2PA specificatie – het voor een (her)gebruiker verifieerbaar maken van de herkomst en authenticiteit van de uiting. Maar een secundaire motivatie kan zijn om hiermee ook naamsvermelding van de maker (technisch) aan de uiting te verbinden. Dit is ook in lijn met het denken over authenticiteit in relatie tot (media)merken, zoals beschreven in het vorige hoofdstuk.[16]

Hierbij moet de kanttekening geplaatst worden dat andere bestaande en breed geaccepteerde standaarden als EXIF, deze informatie ook opslaan. Het verkrijgen van naamsvermelding bij media-uitingen is daarmee niet zozeer een technische uitdaging, maar een vraagstuk over menselijke adoptie en overname door gebruikers.

Generatieve AI tekent zijn eigen product – Een aantal veelgebruikte generatieve AI systemen, met name beeldgeneratoren[17], gebruiken C2PA om generatieve AI te ondertekenen en daarmee te onderscheiden van afbeeldingen uit de fysieke wereld.[18] Ook de opname van AI genereerde onderdelen van een media-uiting kunnen daarmee als assets ondertekend worden.[19] Let op dat het manifest makkelijk verwijderd kan worden, het niet hebben van een C2PA-manifest is geen garantie dat het materiaal door mensen is gemaakt.

Enkele grote mediaplatformen hebben de C2PA specificatie geadopteerd, waaronder Facebook[20], LinkedIn[21], YouTube[22] en het Content Delivery Network CloudFlare[23]. Deze zetten C2PA vooral in om (eigen) gemaakt AI gemaakt materiaal te identificeren.[24] Hiermee kunnen zij de geclaimde

[14] Zie https://lens.truepic.dev/docs/libc2pa-introduction
[15] Bijlage 5 - Beschikbare tools en implementaties
[16] Zie Authenticiteit
[17] Zie OpenAI is adding new watermarks to DALL-E 3 | The Verge
[18] Zie How Google and the C2PA are increasing transparency for gen AI content en Content Credentials in Azure OpenAI
[19] Naast C2PA, zijn er ook alternatieve aanpakken rond GenAI, zoals SynthID van Google. Google is adding AI watermarks to photos manipulated by Magic Editor | The Verge
[20] Zie Labeling AI-Generated Images on Facebook, Instagram and Threads | Meta
[21] Zie Content credentials | LinkedIn Help
[22] Zie Building trust on YouTube: 'Captured with a camera' disclosure
[23] Zie Cloudflare becomes the first major content delivery network to implement Content Credentials
[24] Zie ook Coalitie

herkomst van beelden op een uniforme manier communiceren. Gebruikers van deze social mediaplatformen krijgen in de gebruikersinterface toegang tot een menselijk leesbare variant van (een deel van) het C2PA-manifest. Consumenten kunnen vervolgens in de beoordeling van het werk het vertrouwen dat zij scheppen in de gevalideerde ondertekenaar van het werk meenemen. In theorie is het overigens goed mogelijk dat op hetzelfde sociale mediaplatform twee inhoudelijk tegenstrijdige media-uitingen worden verspreid, waarbij van elk van de berichten de (twee verschillende) omroepen als afzender geverifieerd kunnen worden middels C2PA.

Het is aannemelijk dat een (aanvullende) motivatie voor de ontwikkelaars van dergelijke generatieve AI modellen om hun eigen producten op deze manier te tekenen ook het gebruik van deze informatie voor hun eigen werkprocessen is. Het kan namelijk functioneren als een methode om dergelijke met generatieve AI gegenereerde producten uit te sluiten bij het hertrainen en doorontwikkelen van hun modellen. Dit helpt hen om de uitdaging van entropie (of "model collapse"[25]) tegen te gaan.

De verwerker controleert de echtheid van bronnen – Aan de start van de mediaproductie (of pre-productie, kan de mediamaker, journalist of redacteur, etc. het C2PA-manifest van bronnen controleren op echtheid, om ervoor te zorgen dat er op basis van geverifieerde bronnen verslag wordt gelegd. De verwerker maakt bewust keuzes wat te doen met materiaal dat geen C2PA-manifest heeft, of wanneer dat manifest niet valideert.[26]

De bewerker tekent eigen aanpassingen – De bewerker van het ruwe materiaal waar een media-uiting op gebaseerd is, kan C2PA compatibele software gebruiken om de samenstelling en bewerkingen van een fragment of media-uiting te documenteren en te bevestigen middels een digitale handtekening. Het documenteren van wijzigingen met bewerkingssoftware, zoals het uitsnijden, aanpassen van kleuren, verwijderen van herkenbare gezichten, toevoegen van een audiotrack, etc. kan een goed inzicht in het productieproces en hoe hier met journalistieke integriteit wordt omgegaan geven. Deze use case ziet zich gespiegeld in de suggestie naar meer transparantie van de academici in de interviews.[27]

Wanneer de hard- en software-ondersteuning van C2PA niet (in de gehele productieketen) voorhanden is, kan de media-uiting als geheel ook van een C2PA-manifest voorzien worden. Het is dan aan de laatste maker van het manifest om te bepalen wat het doel van een handmatig toegevoegd manifest is. In dit geval wordt de 'last' van vertrouwen ook sterker bij de eindgebruiker neergelegd.[28]

[25] Zie Model collapse - Wikipedia
[26] Bij de sectie Menselijke factor wordt de wenselijkheid en haalbaarheid van een dergelijke verstrekkende controle door de journalist en redacteur apart besproken.
[27] Zie Menselijke factor en Prof. dr. ir. Jeroen de Ridder en Prof. dr. Yael de Haan
[28] Gegeven de nog zeer beperkte adoptie van C2PA in hardware en softwareproducten, en daartegenover de enorme diversiteit van middelen die voor mediaproductie worden gebruikt, is dit (afhankelijkheid van handmatig toegevoegde manifesten voor media-uitingen) de huidige praktijksituatie.

Distributie

Bij distributie met een C2PA getekende media-uiting kan de ontvanger controleren of het bestand aangepast is tijdens het transport. Wanneer het bestand aangepast is, valideert het C2PA-manifest niet meer. Hiermee kan de ontvanger bijvoorbeeld controleren of een ontvangen PDF, afbeelding of ander document daadwerkelijk van de afzender afkomstig is.

Veel distributiekanalen, waaronder veel social media, transformeren mediabestanden om zo makkelijker, gestandaardiseerde en goedkopere media te verspreiden. Vaak wordt hierbij veel van de metadata, waaronder bijvoorbeeld C2PA-manifest, verwijderd. Het niet hebben van een C2PA-manifest betekent dan niet dat het bestand gemanipuleerd is, maar dat extra aandacht nodig is bij de consumptie van het materiaal.

Databank met soft-bindings – Zelfs wanneer een C2PA-manifest verwijderd is op een distributiekanaal kan via 'soft-bindings' het manifest nog wel teruggehaald worden. Dit vergt dan wel weer dat er een databank met soft-bindings en metadata aangeboden wordt.[29] Een databank met soft-bindings is een databank met digitale handtekeningen en verwijzingen naar de oorspronkelijke bron van de media-uiting. De consument kan deze digitale handtekening achterhalen, zelfs wanneer het werk tot op zekere hoogte is aangepast.

Distributeur biedt C2PA getekend materiaal aan – De distributeur biedt C2PA getekend materiaal aan. De distributeur communiceert actief naar haar gebruikers dat het materiaal getekend is met een C2PA-manifest. Deze distributeurs kunnen zowel content platforms zijn, zoals hoe LinkedIn het Content Credentials (CR) logo in de tijdlijn van gebruikers toont, om (voor nu) AI gegenereerde uitingen aan te merken.[30] Maar dit kan ook de website van de mediamakers zelf zijn, zoals een mediapartij als de CBC/Radio-Canada die op haar website de mogelijkheid biedt om sommige beelden via een link naar ContentCredentials.org te valideren.[31] Zoals eerder gemeld bieden Facebook, YouTube, en LinkedIn dergelijke interfaces al aan.

Consumptie

Bij de consumptie van C2PA getekende content zijn er ten minste drie use cases denkbaar, afhankelijk van de ontvanger: het brede publiek (B2C) en organisaties (B2B).

Organisaties wisselen getekende content uit – Wanneer er in de keten vertrouwd wordt op C2PA, zal iedere partner in de keten C2PA op diens eigen manier valideren en verwerken in haar eigen (journalistieke) werkprocessen. Hierdoor kan, in een divers landschap van producenten en makers, vertrouwd worden op de oorsprong van het werk.

Hier is het van belang om tussen de partijen duidelijk te communiceren wat de intentie is van de C2PA verificatie. Is het een verificatie dat de media zonder misinformatie is? Of is het dat de 'best-effort' van een redactie naar het werk gekeken heeft? Of is het enkel de bevestiging dat de partij afzender van de uiting is?

[29] Zie Bijlage 2 – Technische uitleg C2PA
[30] Zie LinkedIn Adopts C2PA Standard
[31] Zie bijvoorbeeld Content Credentials Verify

Organisaties controleren media-uitingen op GenAI elementen – Met de opkomst van synthetische media, ontstaan of bewerkt met behulp van software en in het bijzonder GenAI modellen voor verschillende type media, kan het inspecteren van een C2PA-manifest (indien aanwezig) helpen bij het identificeren van dergelijke synthetische media-elementen. Dit kan bijvoorbeeld helpen bij het detecteren en signaleren van manipulatieve media-uitingen, bijvoorbeeld in de vorm van vervalsingen en impersonaties.[32] Dit sluit aan op de trend dat diverse GenAI beeldgeneratoren C2PA adopteren om eigen producten te tekenen.[33]

Gebruikers consumeren getekende content – Bij de consumptie door de media-uitingen door het beoogde publiek, is presentatie van cruciaal belang. Dit betekent dat net als bij de bovenbeschreven social media platformen de informatie en doelstelling van de handtekening eenduidig en makkelijk begrijpelijk moet zijn voor de eindgebruiker. Dit laatste punt is makkelijker gezegd dan gedaan, zoals blijkt uit eerder onderzoek (binnen Proof of Provenance[34]) en diverse van onze interviews. Ook hebben verschillende geïnterviewden – zowel vanuit technisch, als vanuit gebruikersperspectief – twijfels geuit over de meerwaarde van de digitale handtekening op zichzelf (als 'vinkje'), voor een consument.[35]

Gebruikers bevestigen elders gevonden (gestript) materiaal – Tenslotte kan een gebruiker ook materiaal terugvinden waarbij ze het vermoeden hebben wie de uitgever is, maar waarbij de metadata (en het manifest) in de distributieketen verloren zijn gegaan. Er ligt hier een kans om deze informatie via soft-bindings en een externe database weer terug te krijgen.

Dit legt de verantwoordelijkheid bij de uitgever om op een centrale plek bij te houden welke media-uitingen door hen verspreid zijn met een C2PA-manifest. Hier komt het vraagstuk over waar macht ligt in dit systeem om de hoek kijken.[36]

Archivering

Aan het einde van de levensloop van de informatie vindt soms archivering van de media-uiting plaats. In het Proof of Provenance project zijn hiervoor al use cases uitgewerkt.[37] Deze worden hieronder samengevat en waar nodig omgezet naar specifieke toepassingen van C2PA.

Metadata extraheren – Archieven beschrijven werken in hun collecties in collectiebeheersystemen, hiervoor zijn ze geïnteresseerd in bijvoorbeeld de naam van de maker, de titel van het werk, de publicatiedatum, etc. Deze informatie is deels te vinden in een C2PA-manifest.

Hertekenen – Na verloop van tijd zijn sommige van de public keys niet meer toegankelijk. Het kan ook gebeuren dat andere onderdelen van validatie niet meer werken, doordat de specificatie niet meer backwards compatibel is. In dat geval moet voor de integriteit van het archiefstuk een

[32] Zie bijvoorbeeld het werk van Witness op dit gebied: https://www.youtube.com/watch?v=VsN7JN0yfCg
[33] Zie Productie
[34] Zie 1.6 Resultaten van Eindrapport IDO21G014 - Proof of Provenance voor communicatie vanuit overheden, erfgoed- en media-instellingen
[35] Zie o.a. Tom Demeyer, Dr. Neal Krawetz and Shawn Masters (English), Prof. dr. Yael de Haan
[36] Zie Coalitie
[37] Zie ook Use case webarchivering die eerder voor Proof of Provenance is opgesteld door Beeld & Geluid

(verifieerbaar) bewijsstuk opgeslagen worden met een bewijs van authenticiteit (en herkomst) van een gearchiveerde media-uiting op het moment van inname. Een archief zal moeten registreren dat het C2PA-manifest kon worden gevalideerd op het moment van inname.

Kritische noten C2PA

De C2PA kent zijn kritieken. Kritiek op de technologie wordt met name door Dr. Neal Krawetz (Hacker Factor) uiteengezet[38], ook andere geïnterviewden (Demeyer en Masters) hebben soortgelijke kritiek geuit. Kritiek op mogelijke implementatie en de menselijke factor hierbij komt o.a. uit de interviews met De Haan.

Vertrouwen komt te voet en gaat te paard

Een door Dr. Krawetz besproken casus van een CBC/Radio-Canada publicatie laat zien hoe precair het vertrouwen dat een C2PA-manifest schept kan zijn.[39] Het betreft hier een foto die het bericht ondersteunt dat er additionele veiligheidsmaatregelen genomen werden bij een moskee in India, rondom nationale verkiezingen.

Uit een analyse van het C2PA-manifest door Dr. Krawetz blijkt dat het manifest geen eenduidige tijdstempels hanteert; deze liggen weken uit elkaar. Hierdoor ontstaat er een vermoeden van manipulatie van de uiting. Ook roept dit de vraag op of het beeld ook daadwerkelijk het bericht ondersteunt?

Bij nadere inspectie blijkt er nog meer conflicterende metadata te zijn: Een meer "originele" versie van de afbeelding werd gevonden (via een directe link, middels reverse image search). Deze bevat EXIF-data van een Sony-camera, terwijl de "geauthenticeerde" versie (in de publicatie) deze metadata ontbeert. Dit maakt het moeilijk om de gebruikte camera of andere opnamedetails te verifiëren.

Tenslotte biedt CBC/Radio-Canada meerdere formaten van de afbeeldingen op de website aan, met en zonder C2PA-manifest. Dit is standaardpraktijk voor online distributie van digitale afbeeldingen, maar het benadrukt ook hoe gemakkelijk C2PA-metadata kunnen worden verwijderd tijdens online publicatie en distributie.

De casus hierboven demonstreert hoe makkelijk er (naar vermoeden, in het geval van CBC/Radio-Canada) tegenstrijdige metadata kan worden opgenomen in een C2PA-manifest. Waarschijnlijk komt dit voort uit gebrekkige implementatie of training. Met hetzelfde gemak kan een kwaadwillende partij dus ook opzettelijk tegenstrijdige metadata in een manifest opnemen.

Technologie

Vervalsing claims – Er werd in 2023 een fundamenteel probleem blootgelegd met de gereedschappen en procedures bij de implementatie C2PA. Hieruit blijkt dat handtekeningen relatief eenvoudig vervalst kunnen worden.[40] Hierdoor kan iemand anders zich relatief eenvoudig voordoen als bijvoorbeeld de NOS. Er is redelijk wat technische kennis vereist om een dergelijke valse claim te weerleggen. Een media-uiting met een vervalste handtekening kan desinformatie

[38] Zie ook het interview met Neal Krawetz en Shawn Masters.
[39] Zie C2PA and the CBC - The Hacker Factor Blog
[40] Zie C2PA's Worst Case Scenario - The Hacker Factor Blog

in de hand spelen. Mohamed Badr Taddist, die de implementatie pilot bij de EBU uitvoerde, heeft bevestigd dat om deze reden self-signed certificaten als 'unknown' worden bestempeld in de veelgebruikte tools.[41] In essentie bepaalt het consortium uit welke bronnen de 'vertrouwde' handtekeningen kunnen komen.

Eigen register certificaten – Het consortium houdt een eigen register bij van Certificaten[42], waardoor handtekeningen die niet onafhankelijk gevalideerd kunnen worden, wel gevalideerd kunnen worden door C2PA tooling, maar niet voorkomen in de gangbare certificate stores. Dit maakt de C2PA kwetsbaar. Gangbare publiek certificates stores staan op elk besturingssysteem, ze zorgen ervoor bijvoorbeeld voor dat browsers HTTPS kunnen gebruiken. Geavanceerde gebruikers kunnen zelf bepalen welke Certificate Authorities deze vertrouwen.[43] In de veelvoorkomende tooling van het consortium kan dit niet.

Uit interviews blijkt dat de gereedschappen om media-uitingen te ondertekenen gebruikmaken van certificaten die Adobe zelf uitgeeft.[44] Doordat deze certificaten buiten de bestaande certificaten-infrastructuren zitten, rust deze keten op de door het consortium aangewezen certificaten. Overige certificaten kunnen gebruikt worden, maar in de tooling van het consortium worden deze aangegeven als 'unknown'. Hierdoor wordt door dienstverlening een centrale positie van macht gecreëerd.

Claims binnen claims – Hoewel de C2PA-specificatie aangeeft dat nieuwe claims de voorgaande claims zouden moeten valideren, is dit geen harde eis. Claims binnen claims worden niet standaard gevalideerd. Wanneer een entiteit een media-item ondertekent, gaan de procedures ervan uit dat de claims die binnen de laatste handtekening voorkomen ook accuraat zijn. Maar in de praktijk wordt enkel gevalideerd dat de laatste handtekening betrouwbaar is. Dit betekent, in combinatie met de eerder genoemde mogelijkheid om handtekening te vervalsen, dat partijen zich kunnen voordoen alsof ze een vertrouwde partij zijn. Dit kan leiden tot de verspreiding van mis- en desinformatie.

Transparantie – Alle geïnterviewde technologen geven aan dat verificatiegereedschappen, zoals beschikbaar op contentcredentials.org, niet compleet transparant zijn. Op het moment van verificatie is een certificaat van het consortium nodig. Het is niet helemaal zeker of hier geen lek van persoonsgegevens kan ontstaan (in de toekomst). De beschikbaar gestelde gereedschappen zouden ingezet kunnen worden om een profiel op te stellen van de mediaconsument. Er is stevige controle nodig dat deze tools geen persoonsinformatie lekken. Een groot deel van de coalitie heeft namelijk een bedrijfstak gericht op (persoonlijk gerichte) advertenties. In deze context wijst De Haan op de gevoeligheid van journalisten voor dergelijke risico's bij de adoptie van technologie.[45]

Digitale identiteit – Demeyer geeft daarnaast aan dat C2PA, in tegenstelling tot Proof of Provenance, geen verplichting heeft om te tekenen met een digitale identiteit (destijds middels IRMA). Hierbij wordt een derde autoriteit gebruikt om te bevestigen dat jij inderdaad de persoon bent die ondertekent. Demeyer geeft bijvoorbeeld aan dat het in Photoshop mogelijk is om

[41] Zie Lucille Verbaere em Mohamed Badr Taddist (EBU) (English)
[42] Zie C2PA and Untrusted Certificates - The Hacker Factor Blog
[43] Zie "5.3.2. Use of the Private Credential Store" in C2PA Implementation Guidance
[44] Zie C2PA and Untrusted Certificates - The Hacker Factor Blog
[45] Zie Prof. dr. Yael de Haan

"NOS" als naam op te geven en deze te ondertekenen.[46] Er is geen verplichte controle op de naam van de tekenaar. In voorgaande versies van C2PA was het nog mogelijk om een digitale identiteit toe te voegen (W3C Verifiable Credentials), deze is in de 2.0 versie verwijderd ten faveure van eigen nog te ontwikkelen extensie van C2PA: een identity assertion van de Creator Assertions Working Group.[47] Deze wordt nog opgesteld, en is daarmee out-of-scope voor deze analyse.

"glass-to-glass" ambitie – C2PA heeft de ambitie om van "glass-to-glass"[48] inzicht te geven in de herkomst en authenticiteit van een media-uiting. In de praktijk bestaat elke media-uiting namelijk uit een serie van aanpassingen en bewerkingen, die tot het uiteindelijke resultaat dat gereed is voor distributie leiden.

Een complete implementatie van C2PA in het productieproces van een media-uiting vergt dat de hardware en software aangeschaft worden om deze stappen goed te kunnen documenteren. Met andere woorden: De technische middelen waarmee de media-uitingen tot stand komen moeten *allemaal* C2PA (correct) ondersteunen. Bijlage 5 laat zien dat de beschikbare implementatie in soft- en hardware nog beperkt is. De auteurs van dit rapport betwijfel ook ten zeerste of een benadering van deze ideale situatie realistisch is.

Coalitie en governance

Big tech – De C2PA coalitie bestaat onder andere uit grote techbedrijven als Adobe, Amazon, Google, Intel, Meta, Microsoft, OpenAI, Sony en TruePic. Deze coalitie biedt gereedschappen aan voor verifieerbare provenance en authenticiteit. Hierdoor ontstaat ook een machtsrelatie. Het is verstandig om deze machtsrelatie uit te lichten. Deze partijen hebben profijt van deze controle over gereedschap en infrastructuur. De reden om deze technologie te ontwikkelen, te promoten en te ondersteunen, hoeft niet gelijk te zijn aan de doelstelling van mediapartijen om deze technologie te gebruiken. In de woorden van Michel Foucault:

> *"We are subjected to the production of truth through power; and we cannot exercise power except through the production of truth."*[49]

Tegelijkertijd kan macht ook uitgeoefend worden door andere spelers in de distributieketen van media-uitingen (die voorzien zijn van C2PA-manifesten). Distributeurs kunnen bijvoorbeeld de manifesten (selectief) verwijderen, waardoor de illusie ontstaat dat er een verschil is tussen media-uitingen van verschillende afzenders. Ook kan een nieuwsorganisatie die C2PA gebruikt haar geloofwaardigheid proberen te versterken met behulp van authenticiteit en zo meer invloed uitoefenen op de publieke opinie.

In een ideale wereld ligt de autoriteit bij onafhankelijke partijen, zodat er geen één autoriteit aan de structuren kan morrelen. Er zijn garanties nodig voor de afscherming van de data en de

[46] Zie Tom Demeyer
[47] Identity Assertion
[48] Een term uit de mediaproductie die duidt op het gehele werkproces van de lens (glass) waarmee de opname wordt gemaakt tot aan het beeldscherm (glass) waarop de uiting wordt bekeken.
[49] Michel Foucault, Michel, Power/Knowledge: Selected Interviews and Other Writings. (Edited by Colin Gordon. New York: Pantheon, 1980), 131, 133 and 93 respectively.

onmogelijkheid van manipulatie. Hiervoor is transparantie nodig, in de specificatie, in de implementatie en in het beheer van dit geheel.[50]

Om te kunnen stemmen op de richting en inhoud van de specificatie is lidmaatschap vereist. Dit brengt kosten met zich mee (een jaarlijks bedrag van €5.000).[51] Lidmaatschap van het consortium wordt gedomineerd door de big-tech bedrijven waardoor ook de beslissingen in het belang van deze organisaties gemaakt zullen worden.

Generative AI – De geïnterviewden[52] zien dat C2PA ook ingezet kan worden om onderscheid te maken tussen AI-gegenereerde beelden en mens-gemaakte beelden[53]. Hierbij wordt door Krawetz en Masters gesuggereerd dat de C2PA-partners die AI trainen, de technologie inzetten om producten van AI te filteren uit trainingssets, als maatregel tegen "model collapse"[54]. Voor de eindgebruiker is deze additionele functionaliteit van de standaard interessant. Er moet echter wel rekening mee gehouden worden dat het gebrek aan een C2PA-manifest niet betekent dat het niet gaat om AI-gegenereerd materiaal.[55]

Belangenverstrengeling – In het interview met Dr. Neal Krawetz and Shawn Masters worden zorgen geuit over mogelijke belangenverstrengeling van Leonard Rosenthol (Senior Principal Scientist with Adobe Systems and chief architect of C2PA). Rosenthol leidt namelijk tevens het reviewproces om van C2PA een ISO standaard te maken. De kritiek van Dr. Neal Krawetz en Shawn Masters is dat de slager hier zijn eigen vlees keurt.

Mediasector (omroepen) – In de mediasector zijn diverse voorbeelden van adoptie van de specificatie te vinden, bijvoorbeeld bij de EBU, CBC/Radio-Canada, WDR[56] en BBC[57]. Maar van brede adoptie is vooralsnog geen sprake. Op geen van de hoofdplatformen is een media-uiting met een C2PA manifest vindbaar. De EBU heeft de werkgroep "C2PA Workflows" opgericht, met als doel om de ontwikkeling en adoptie van C2PA productie- en distributietechnologie te ondersteunen.[58] Als onderdeel van dit onderzoek hebben we gesproken met de bij de EBU betrokken personen. De hoofdontwikkelaar van de implementatie van de EBU zelf, een demonstrator voor de IBC conferentie[59], is stagiair bij de EBU.[60]

---

[50] Zie ook interview prof. dr. Ir. de Ridder
[51] Zie Membership - C2PA
[52] Zowel Dr. Neal Krawetz and Shawn Masters als de EBU
[53] Dit sluit tevens aan op een recente wending in de positionering van C2PA: Van nadruk op de mogelijkheid om middels C2PA uitingen als afzender te ondertekenen en daarmee van authenticiteit te voorzien, naar nadruk op de mogelijkheid voor consumenten of hergebruikers om middels C2PA uitingen te herkennen die het product van generatieve AI zijn.
[54] Zie Model collapse - Wikipedia
[55] Zie "Weaknesses" op pagina 17 van SWOT and GAP Analysis of the Proof of Provenance project – PublicSpaces.net
[56] WDR erprobt KI-Siegel: "Digitaler Beipackzettel" für mehr Transparenz - Presselounge
[57] Mark the good stuff: Content provenance and the fight against disinformation en Embedding transparency, enhancing trust
[58] Zie C2PA workflows | EBU Technology & Innovation
[59] Zie EBU showcases multi-partner end-to-end C2PA workflow at IBC | EBU Technology & Innovation
[60] Zie Lucille Verbaere and Mohamed Badr Taddist (EBU) (English)

Werkprocessen

C2PA is een middel om de authenticiteit van provenance informatie te communiceren. Het is een technologisch middel en moet dan ook correct en voor de juiste toepassing ingezet worden. De hele keten van mediamakers moet deze gereedschappen op de juiste manier inzetten om mogelijke fouten te voorkomen.[61] Wanneer dit niet goed gebeurt, stelt de mediapartij zich (waarschijnlijk) bloot aan reputatieschade.

Potentiële reputatieschade – Bronnen, mediamakers, en distributeurs kunnen – bewust of onbewust – C2PA-manifesten gebruiken om ogenschijnlijk betrouwbare informatie, maar feitelijk onjuiste informatie, in de contentketen te introduceren.

Om dit te vermijden is het aan te bevelen om alle processen van het verkrijgen van media, media aanpassen en media distribueren goed door te lichten en procedures op te stellen hoe met C2PA om te gaan. Communiceer deze procedures ook publiekelijk, zodat de gebruiker en mediaconsument weet wat voor soort authenticiteit er met C2PA bedoeld wordt. Dit helpt de kans verminderen dat authenticiteit en waarheid met elkaar worden verward.

Impact van implementatie op de werkvloer – Bij de invoering van C2PA is het van belang om te denken aan de dagelijkse impact van de standaard. Complexe administratieve handelingen en controles zullen moeilijk te adopteren zijn in de mediapraktijk, waar tijdsdruk en beperkte middelen gemeengoed zijn. Dit ideaal van frictieloos is nastrevenswaardig. Het is wel de vraag of deze haalbaar is, zonder compromis. Het C2PA-manifest vergt administratieve handelingen. Wanneer deze niet automatisch door de ingezette gereedschappen gedaan worden dan is enige frictie onvermijdbaar. Het is ook nog eens de vraag of een degelijke implementatie zonder (steekproefsgewijze) menselijke controle van de manifesten mogelijk is gezien de gelaagdheid en complexiteit van de manifesten en bijvoorbeeld ook het risico dat anonieme bijdragers bij distributie worden blootgegeven via het C2PA-manifest.

Hard- en software & training – De ideale C2PA implementatie "from glass to glass", waar de specificatie vanuit gaat, vergt niet alleen een brede technische adoptie. Er zullen ook vormen van opleiding en handhaving van de gebruikers van deze hardware en software nodig zijn, als het gaat om het (correct) toepassen, vullen en tekenen van de C2PA-manifesten. Met andere woorden: Het moet voor de gebruikers duidelijk zijn met welk doel de C2PA-manifesten worden aangemaakt en wat hier de betekenis van is. En dit alles moet uiteraard ook gebruiksvriendelijk en efficiënt zijn. En vervolgens op de juiste manier naar de eindgebruikers/consumenten van deze informatie gecommuniceerd worden.

Eindgebruiker

De Code Journalistiek Handelen[62] stelt kwalitatieve criteria aan de processen en procedures van de journalistiek. Geïnterviewden verwachten daarom dat digitaal ondertekenen van uitingen maar weinig toegevoegde waarde levert. De kwaliteit van het merk zelf is belangrijker dan een

[61] Zie CBC/Radio-Canada case: C2PA and the CBC - The Hacker Factor Blog
[62] Code Journalistiek Handelen

technische bevestiging van de authenticiteit van herkomstinformatie. Bestrijden van des- en misinformatie ligt volgens de Haan ogen meer bij de burger.

Deze zienswijze wordt onderschreven door Prof. dr. ir. de Ridder[63]. Hij geeft in eerste instantie aan dat het overdreven is dat we omkomen in desinformatie. Prof. Dr. de Haan[64] geeft aan dat er geen hard bewijs is dat mensen onvoldoende informatiebronnen tot zich nemen. De Ridder gelooft wel aan dat een bevestiging van de herkomstinformatie enig vertrouwen kan terugwinnen in de gevestigde media. Maar stelt dat dit er tegelijkertijd ook voor kan zorgen dat deze media meer effectief en bewust worden uitgesloten.

Transparantie – de Ridder wijst op het belang van transparantie in het proces en maakt daarbij een vergelijking met open science, waarbij transparantie van bronnen en processen onderdeel zijn van de wetenschappelijke methode. Tegelijkertijd stelt hij vraagtekens bij de haalbaarheid van een dergelijke werkwijze voor journalistieke uitingen. Het hele journalistieke proces vastleggen, zoals bij open science, is waarschijnlijk te tijdrovend binnen de journalistieke praktijk. [65] Dit wordt verder onderschreven door De Haan.

Beiden zien tegelijkertijd ook de meerwaarde van het bieden van meer transparantie (middels C2PA). De Haan ziet de meerwaarde van C2PA vooral in de bewustwording die dit bij de journalist afdwingt in het vastleggen en tekenen van herkomst. Hierdoor worden bewuste checks ingebouwd in de bestaande workflows. In het verlengde zou in deze vorm C2PA voor de consument interessant kunnen zijn als vorm van transparantie: een keten van herkomst en checks die door de journalist op verschillende momenten in de totstandkoming van de uitging zijn toegepast[66]. In C2PA als slechts een 'vinkje' ziet zij veel minder. Een visie die onder andere door Demeyer gedeeld wordt.[67]

Realistisch gezien ligt het laaghangende fruit bij de uitleg over welke principes worden gehanteerd en waar men met klachten terecht kan. Voor vertrouwen is de intermenselijke relatie in de praktijk ook van belang: "Wat je kent, vertrouw je meer."[68]

Gestandaardiseerde transparantie

C2PA geeft een specificatie hoe met transparantie omgegaan kan worden. De specificatie beschrijft velden als versiegeschiedenis van de media-uiting, tekendatum, creatiedatum, etc. Deze velden scheppen een verwachting van transparantie. Wanneer deze stappen niet rigide gevolgd worden is de kans op reputatieschade groot.

Zo bestaat er een casus over de verificatie van de BBC van een aangepast filmpje.[69] In deze casus heeft de BBC een vervalste video van een C2PA certificaat voorzien. De video, een

[63] Zie Prof. dr. ir. Jeroen de Ridder
[64] Zie Prof. dr. Yael de Haan
[65] Zie Prof. dr. ir. Jeroen de Ridder
[66] Zie Prof. dr. Yael de Haan
[67] Zie Tom Demeyer
[68] Zie Prof. dr. ir. Jeroen de Ridder
[69] Zie IEEE, BBC, and C2PA - The Hacker Factor Blog

montage van twee social media video's met toegevoegde tekst en logo's, was geverifieerd door de factchecking unit van de BBC (BBC Verify).

Uit een analyse van dr. Krawetz bleek dat de BBC drie verschillende manifesten voor dezelfde video had gepubliceerd. Een van de video's, die al was geauthenticeerd, bleek tijdens de productie aangepast te zijn. Een van de twee social media video's was voordat het bij de BBC kwam bewerkt met een toegevoegde soundtrack (luide geweerschoten) om een dramatischer effect te creëren. De factcheckers van de BBC hadden dit niet opgemerkt.

Vervolgens heeft de BBC de video aangepast, ze hebben de geweerschoten eruit gehaald en het geheel van een nieuw certificaat voorzien. Dit certificaat laat echter niet zien dat het werk aangepast is. Deze aanpassing wordt buiten het C2PA-manifest beschreven.[70] Daarbij is er een nieuwe fout in het manifest geslopen. De media-uiting van de BBC lijkt nu uitgeven te zijn (4 maart 2024) voordat de video gemaakt is (6 maart 2024).

Deze casus laat zien dat niet alleen de inhoud van de video, maar het gebrek aan juiste informatie over redactioneel proces tot reputatieschade kan leiden.

Mitigerende stappen

Er zijn enkele mitigerende maatregelen die het consortium kan treffen om antwoord te geven op deze kritische noten.

Digitale identiteit verplichten – Het ontbreekt in de specificatie aan het kunnen toevoegen van een digitale identiteit. Hierdoor kan de specificatie wel ingezet worden om te bevestigen dat een uiting authentiek is, in de zin van onveranderd, maar kan niet bevestigd worden dat de uiting authentiek afkomstig is van een verifieerbare entiteit. Zorg ervoor dat digitale identiteiten een (verplicht) onderdeel van de standaard worden en maak dit vervolgens in elk geval een verplicht onderdeel van de adoptie van C2PA door de NPO.

Certificaten stores – Zorg ervoor dat het consortium en de beschikbare tools gebruik maken van en helder communiceren welke root certificate stores gebruikt worden, als vertrouwde autoriteiten voor de validatie. Zorg ervoor dat de gebruiker weet welke root stores beschikbaar zijn in welke tools en welke impact dit kan hebben op de verificatie.

Kritische participatie – Kritische noten op de coalitie zijn moeilijker te beantwoorden. De richting en inhoud van de specificatie wordt met name bepaald door organisaties (technologische partijen) die niet noodzakelijk dezelfde doelstellingen hebben als de publieke omroepen. Wordt betalend lid van het consortium en blijf kritisch op voorstellen en hou scherp in de gaten of C2PA fit-for-purpose blijft. Zet hierbij parallel in op actieve (en kritische) participatie in de werkgroep van de EBU en zie erop toe dat aanbevelingen vanuit de gemeenschap ook leiden tot de noodzakelijke aanpassingen in de specificatie en het omringende instrumentarium zelf.

Informeer en controleer – Mensen zijn de zwakke schakel bij het gebruik van C2PA. C2PA niet volledig of verkeerd gebruiken kan leiden tot reputatieschade, of een verlies van vertrouwen. Informeer en controleer het gebruik van C2PA gedurende het gebruik van de specificatie.

[70] Zie The Hacker Factor Blog- C2PA and Authentication Updates

Ditzelfde geldt ook voor eindgebruikers. Eindgebruikers die de implementatie van C2PA verkeerd begrijpen kunnen ook reputatieschade veroorzaken.

Bevindingen – La Trahison des images

Het beroemde schilderij van René Magritte 'La Trahison des images' beter bekend door het onderschrift 'Ceci n'est pas une pipe' beschrijft de relatie tussen een fysieke pijp en een afbeelding van de pijp. De afgebeelde pijp is daarmee vooral dat, geen échte pijp. De kijker wordt gewezen op de illusie van representatie. Het is een reflectie op het gebruik van taal, die ons dwingt kritisch na te denken over wat we zien en hoe we dat benoemen.

Figuur 3. Parodie op 'La Trahison des images' door René Magritte, genereerd door Google Gemini.

Ook bij C2PA is deze relatie voelbaar. De specificatie van de *Coalition for Content Provenance and Authenticity* is ter bevordering van content provenance en authenticiteit. Bij het kritisch lezen van de technische specificatie heeft C2PA meer gemeen met de afgebeelde pijp dan met de échte pijp.

Daarmee blijft de C2PA-specificatie alsnog een waardevol instrument, maar het is geen wondermiddel tegen desinformatie. Het biedt geen garantie voor de daadwerkelijke herkomst en daadwerkelijke authenticiteit. Laat staan dat het iets zegt over waarheid.[71] De specificatie beschrijft *hoe* herkomst en authenticiteit vastgelegd en gecontroleerd kunnen worden, maar het garandeert niet de feitelijke herkomst of waarheid van deze informatie. Mits goed geïmplementeerd en aan de gebruiker juist geïntroduceerd is het een instrument voor de gebruiker om beter geïnformeerde beslissingen te nemen. Het is daarmee, net als de pijp van Magritte, een nuttige en waardevolle representatie.

Aanbevelingen

Is C2PA geschikt om misinformatie en desinformatie te voorkomen? Hier kan ten dele een positief antwoord op worden gegeven. C2PA is een gereedschap dat ingezet kan worden om uitingen te voorzien van verifieerbare herkomstinformatie. Of het hiermee des- en misinformatie kan voorkomen, hangt echter af van de inzet van C2PA.

Wanneer een mediapartij ervoor kiest om C2PA voor de bestrijding van misinformatie en desinformatie in te zetten, dan zijn de volgende aanbevelingen van toepassing:

Verstevig de specificatie – In het hoofdstuk met kritische noten is een aantal mitigerende maatregelen ten aanzien van de huidige kritiek op de C2PA specificatie beschreven.[72] Het is verstandig dat de mediapartij(en)zich inzet(ten) om de specificatie in lijn met deze maatregelen te verstevigen wanneer ze ervoor kiezen om C2PA in te zetten.

[71] Met name dat laatste erkent C2PA zelf ook ruimhartig in de specificatie: "C2PA specifications SHOULD NOT provide value judgments about whether a given set of provenance data is 'good' or 'bad,' merely whether the assertions included within can be validated as associated with the underlying asset, correctly formed, and free from tampering." Zie Content Credentials : C2PA Technical Specification
[72] Zie: Kritische noten C2PA

Bepaal de doelstelling – Voorafgaand aan de implementatie van C2PA, is eerst een duidelijke en beperkte scope voor de inzet van C2PA nodig. Bespreek met redacties hoe en waarvoor C2PA ingezet gaat worden. Beschrijf deze doelstelling nauwkeurig en beknopt.

Begin klein, teken alles – Dit lijkt contra-intuïtief, maar begin klein door bijvoorbeeld in eerste instantie alles op een officiële mediaserver van de mediapartij automatisch te tekenen (als een proces tussen content management systeem en publicatieserver). Dit in de plaats van de introductie van C2PA in de gehele keten. Denk hierbij concreet aan het (geautomatiseerd) teken van media-uitingen die via de officiële distributiekanalen van de mediapartij worden verspreid, zoals een video on demand dienst en het CMS van de mediapartij of programmawebsites.

Communiceer – Zorg ervoor dat doelstelling voor zowel medewerkers als consumenten helder en eenduidig te begrijpen is. Bijvoorbeeld: "Deze handtekening laat verifieerbaar zien dat omroep/programma x de uiting zoals je hem nu ziet officieel erkent als afzender en dat gangbare interne procedures professioneel zijn doorlopen." Let op met het koppelen van een bewijs van de afzender aan een claim van 'objectiviteit' van die partij/dat 'merk'.

Introduceer een eigen vorm van verificatie – Er is een risico dat gebruiksinformatie lekt naar de hosters van de verificatiemiddelen (zoals contentcredentials.org). Zorg ervoor dat de mediapartij zelf een verificatietool beschikbaar stelt die hier expliciet op let. Dit stelt de mediapartij ook in staat om een eigen 'trust list' te bepalen voor deze verificatietool.

Licht de processen door – Licht de processen van het verkrijgen van media, media aanpassen en media distribueren goed door. Stel procedures op hoe met C2PA omgegaan kan (en moet) worden. Dit zorgt voor doelgerichte en doordachte adoptie van C2PA in de processen zelf.

Communiceer helder (in lijn met de doelstelling) – Er moet helder gecommuniceerd worden wat de intentie van het C2PA-manifest is voor de mediapartij en partners. Een ondertekend document betekent niet dat de keten van creatie tot productie onder de controle van de uitgevende partij is. Er kan in die keten nog van alles misgaan met het manifest, door het transport, door fout geconfigureerde software, door een menselijke fout, door externe distributiekanalen, etc.

Verhoog transparantie – Ongeacht de inzet van C2PA is het aan te bevelen om procedures die gebruikt worden om authenticiteit van uitingen te verifiëren ook publiek en transparant te delen, zodat de gebruiker weet wat de omroepen doen om bronnen te verifiëren.

Overweeg SEAL – C2PA is niet de enige standaard, hoewel SEAL[73] vrijwel geen adoptie kent is het alsnog het overwegen waard om media-uitingen op basis van een internetdomein (bijvoorbeeld npo.nl) te tekenen. Deze gereduceerde strikte methode helpt bij het aangeven dat er gedurende transport (publicatie) het bestand onveranderd is gebleven.

[73] Zie Bijlage 4 – Secure Evidence Attribution Label (SEAL)

Bijlagen

- Bijlage 1 – Samenvatting Proof of Provenance (PoP)
- Bijlage 2 – Technische uitleg C2PA en vergelijking met SWOT van PoP
- Bijlage 3 – Interviews
    - Tom Demeyer
    - Dr. Neal Krawetz and Shawn Masters (English)
    - Lucille Verbaere and Mohamed Badr Taddist (EBU) (English)
    - Prof. dr. Yael de Haan
    - Prof. dr. ir. Jeroen de Ridder
- Bijlage 4 – Secure Evidence Attribution Label (SEAL)
- Bijlage 5 - Beschikbare tools en implementaties

Bijlage 1 – Samenvatting Proof of Provenance (PoP)

"Proof of Provenance" (PoP) was een project in 2022 dat een middel ontwikkelde voor instellingen en personen om een certificaat van herkomst te creëren om aan te geven dat een persoon met bepaalde kenmerken (attributen) instaat voor de echtheid van een media-uiting. Het project heeft hiervoor een proof of concept opgeleverd, inclusief onderzoek naar mogelijke vervolgstappen. Parallel aan de ontwikkeling van dit proof of concept is een marktonderzoek uitgevoerd door IP Squared, bestaand uit een SWOT van alternatieve oplossingen ten opzichte van PoP. Een van de destijds bekeken alternatieven was C2PA.

Sinds het SWOT rapport uit juni 2022 van het POP project is vooral C2PA significant verder ontwikkeld. Zo zijn Google, OpenAI, Meta en Amazon toegetreden aan het consortium. Ook de specificatie is geüpdatet[74] sinds 2022 van versie 1.0 naar versie 2.1 waaronder veel technische aanpassingen en introductie van ondersteuning voor meer soorten bestandsformaten (EPUB, OOXML, ODF, OpenXPS). Er is inmiddels ondersteuning in verschillende officiële versies van Adobe producten (Photoshop, Lightroom en Firefly) voor alle gebruikers van Creative Cloud.[75] Met de integratie in LinkedIn heeft het tonen van C2PA informatie aan eindgebruikers een belangrijke stap gemaakt.[76] In 2025 is Cloudflare ook toegetreden, gebruikers van Cloudflare kunnen nu selecteren dat afbeeldingen die door het CDN gaan niet gestript worden van deze metadata.[77] Kijkend naar het mediaveld, dan zijn de experimenten die vanuit de EBU worden uitgevoerd met vroege implementaties van de specificatie noemenswaardig.[78]

In de Strengths, Weaknesses, Opportunities and Threats (SWOT) analyse[79] kwam uit het PoP project de volgende matrix naar voren (**vet** gedrukt toegevoegd voor dit rapport):

Strengths	Weaknesses
What are PoP's defining strengths in relation to similar initiatives? ● Open specification ● Public institutions ● Sufficient technological expertise	What are PoP's defining weaknesses in relation to similar initiatives? ● No simple user experience ● Not available on social media platforms ● Attached metadata ● False negatives, unsigned expressions
Opportunities	**Threats**
Given the similar initiatives, what are possible opportunities to mitigate the weaknesses? ● Storing metadata (certificates) on an external (distributed) service ● **Aligning with other initiatives like**	Given the similar initiatives and PoP's strengths, what are remaining threats to PoP: ● External temporary funding ● Low adoption ● Technological vendor lock-in

[74] Zie Content Credentials : C2PA Technical Specification
[75] Zie Content Credentials
[76] Zie LinkedIn Adopts C2PA Standard
[77] Zie Cloudflare becomes the first major content delivery network to implement Content Credentials
[78] Zie EBU joins early implementers of new standard to track media content provenance
[79] "SWOT and GAP Analysis of the Proof of Provenance project", IP Squared, juni 2022

C2PA and MOVA ● Further open up the PoP consortium ● Heritage institutions as providers of provenance and authentication services	

Figuur 4. SWOT Matrix, overgenomen uit
"SWOT and GAP Analysis of the Proof of Provenance project", IP Squared, juni 2022

Vult C2PA de GAP van Proof of Provenance?

Kijkende naar het geheel van deze analyse is de volgende reflectie op strengths, weaknesses en threats te maken. De opportunities worden niet vergeleken, omdat deze te specifiek aan PoP zijn.

Strengths – de sterke punten van PoP komen overeen met de sterke punten van C2PA.

Weaknesses – Door de grotere adoptie is de beschikbaarheid van C2PA al enigszins aanwezig op sociale media, waaronder LinkedIn.[80] De gebruikersinterface lijkt beter ten opzichte van de proof of concept van PoP. Ook biedt C2PA mogelijkheden om manifesten buiten de asset op te slaan.

De zwakte van fout-negatieven is een fundamenteel media probleem, waar C2PA ook geen oplossing voor biedt.

Opportunities – Niet relevant om deze te vergelijken. Een kans voor de NPO is om aan C2PA bij te dragen door bijvoorbeeld een contentviewer te ontwikkelen voor distributieplatformen van de NPO en de omroepen, zodat consumenten (de relevantie informatie uit) de C2PA-manifesten direct op deze omgevingen kunnen inspecteren. Denk hierbij bijvoorbeeld aan NPO Start en het CMS voor omroep- en programmawebsites.

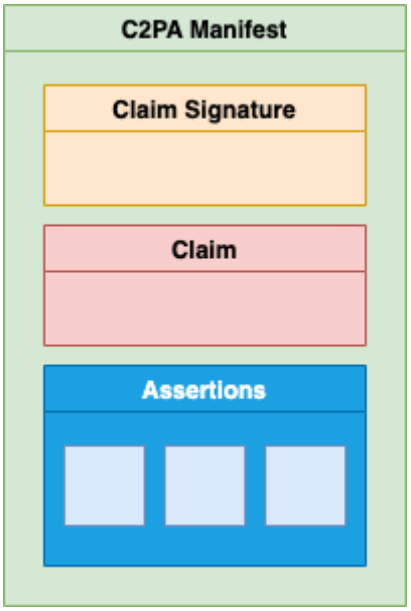
Threats – De financiering en de adoptie die mogelijk bedreiging waren bij de implementatie van PoP zijn niet meer relevant bij C2PA. Door de open specificatie is vendor lock-in waarschijnlijk ook geen bedreiging.

Al met al ziet het ernaar uit dat C2PA een betere kandidaat is dan Proof of Provenance voor de doelstelling van bewijzen van herkomstinformatie toe te voegen aan uitingen van mediabedrijven.[81] Hierbij dienen de mitigerende stappen die in het hoofdstuk Kritische noten C2PA worden beschreven wel in acht genomen te worden.

[80] Zie https://news.linkedin.com/2024/May/linkedin-rolls-out-c2pa-ai-generated-content-standard
[81] Dat wil niet zeggen dat C2PA niet te bekritiseren is. Zie Kritische noten C2PA

Bijlage 2 – Technische uitleg C2PA

C2PA is ontworpen om wereldwijd, opt-in en interoperabel te zijn. Er zijn een breed aantal hardware- en software implementaties beschikbaar voor verschillende soorten mediaformaten. Het opslaan van herkomstinformatie wordt 'Content Credentials' genoemd. Deze credentials worden opgeslagen in een manifest. Een C2PA-manifest bevat metadata, informatie over de herkomst van een (deel) van het beschreven mediabestand (een asset). De informatie over een geheel mediabestand kan in het bestand zelf of extern opgeslagen worden.



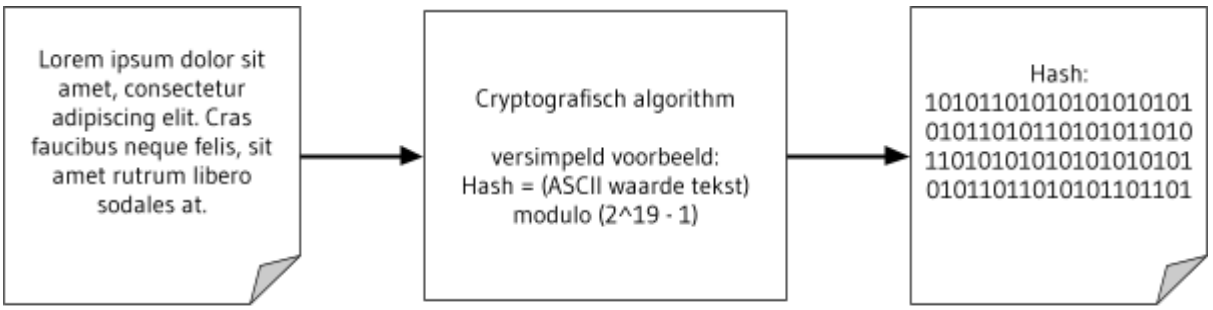
Figuur 5. grafische weergave C2PA-manifest

Een C2PA-manifest bestaat uit beweringen, verklaringen, en handtekeningen:

- **Assertions (beweringen) –** een bewering heeft betrekking tot (een deel van) het mediabestand (de asset). Waaronder bijvoorbeeld beweringen over het auteursschap, maakproces, bewerkingen en de gebruikte technologie.
- **Claims (verklaringen) –** beweringen worden aangevuld met informatie over de entiteit die deze informatie bevestigt.
- **Signature (handtekening)** – Een claim wordt voorzien van een digitale handtekening. Deze bevestigt de bewering, dat betekent overigens niet dat de bewering juist is.

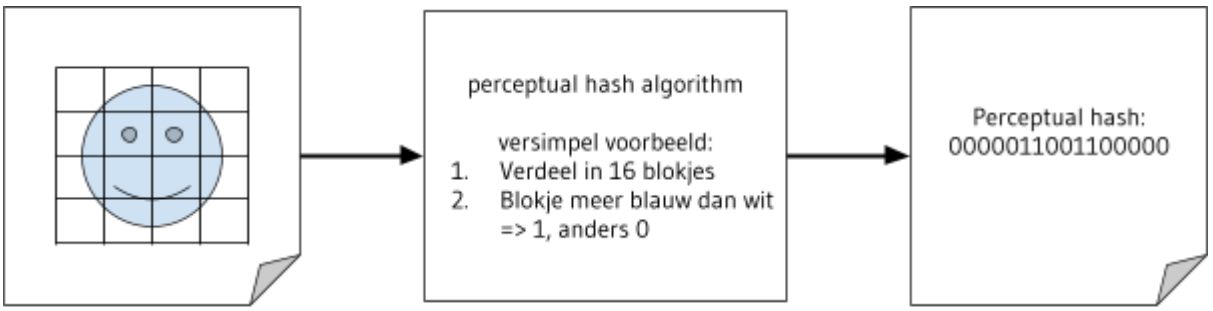
De digitale handtekening op de claim wordt met behulp van een privésleutel van een ondertekenaar getekend. Hierbij wordt de veelgebruikte X.509 public key standaard gebruikt. De identiteit van de ondertekenaar die aan de cryptografische ondertekeningssleutel is gekoppeld, is de basis van vertrouwen in C2PA.

Om assertions en claims over de (deel van het) mediabestand aan een bestand te koppelen kunnen twee soorten verbindingen gemaakt worden (Content Bindings). Een manifest kan meerdere van deze bindings bevatten (als assertion).



Figuur 5. Voorbeeld werking cryptografische hash.

- **Hard bindings** – cryptografische handtekeningen (hashes) die de gehele asset of een deel daarvan uniek identificeren. Wanneer een bestand ook maar met 1 bit aangepast is zal de binding niet meer valideren. Er is maar 1 hard binding per manifest mogelijk.



Figuur 6. Voorbeeld werking perceptual hash.

- **Soft bindings** – Perceptual hashing of soft bindings kijken naar hoe de informatie in een bestand herkenbaar is. Zo kan het ook aanpassingen aan een bestand overleven. Denk aan het omzetten van het bestandsformaat, het embedden van bestanden en het verkleinen van het bestand. Soft bindings zijn gebaseerd op de inhoud van het media bestand, en niet op het bestand zelf. Ze kunnen ook na aanpassingen op het bestand valideerbaar zijn.

Er zijn meerdere validaties op een C2PA-manifest mogelijk. De validatie kan betekenen dat de entiteit die getekend heeft dit deed op een moment dat het certificaat geldig was. Er kan gecontroleerd worden of de handtekening ook van toepassing is op het daadwerkelijke bestand, door de soft en hard bindings.

De C2PA heeft richtlijnen die ernaar streven dat de specificatie de privacy en persoonlijke controle over gegevens respecteert, met een kritisch oog voor potentieel misbruik.

Bijlage 3 – Interviews

Deze bijlage bevat samenvattingen van diepte-interviews met relevante denkers en doeners in het veld van authenticiteit, journalistiek, des- en misinformatie. Deze staan hieronder in chronologische volgorde waarin ze zijn geïnterviewd. De interviews vonden plaats in de periode februari-maart 2025.

De verslagen zijn na het interview voorgelegd aan de geïnterviewden met de vraag of dit een accurate representatie is van het gesprek. De samenvatting die opgenomen is, is als gevolg van het voorgaande met goedkeuring van de geïnterviewden opgenomen. De interviews met internationale partijen zijn in het Engels uitgevoerd en samengevat, dit om de validatie van de samenvatting nadien te vergemakkelijken.

Tom Demeyer

Tom Demeyer was CTO van Waag Society gedurende het Proof of Provenance project. Demeyer is expert op het gebied van data-analyse, data-ethiek, privacy, op veiligheid gerichte technologieën en het semantisch web. Hij werkte tevens mee aan digitale identiteit bij de Gemeente Amsterdam en het Ministerie van Binnenlandse Zaken.[82]

Vanuit zijn betrokkenheid bij Proof of Provenance project, heeft Demeyer de C2PA als een complementair initiatief gezien, dat verder kan bijdragen aan het vaststellen van de authenticiteit van media-uitingen in het huidige landschap. Dit "text-based manifest" zou bijvoorbeeld meegenomen kunnen worden in een handtekening die met de Proof of Provenance technologie gezet wordt.

Deze houding van Demeyer komt voort uit een breder geloof in de meerwaarde van cryptografie, mits deze technieken goed geïmplementeerd worden (in specificatie en toepassingen). C2PA heeft daarmee de potentie om bij te dragen aan een duidelijker medialandschap (als het gaat om authenticiteit van uitingen). Of deze belofte wordt ingevuld moet blijken uit de ontwikkeling, implementatie en adoptie van de standaard.

Tegelijkertijd zijn er wel een aantal kanttekeningen die Demeyer bij de specifieke aanpak van C2PA maakt:

- Hij maakt zich zorgen over de toegankelijkheid van de technologie en standaard voor kleine makers, de uitstraling en adoptie is momenteel nog erg 'corporate'
- Hij vind het jammer dat - anders dan bij Proof of Provenance - er geen directe koppeling met een digitale identiteit/wallet is zoals Yivi (voormalig IRMA)[83], AID[84] of EU Digital Identity[85]. De meeste mensen hebben geen eigen x.509 certificaat. En het is onduidelijk of anoniem tekenen mogelijk is, of dat dit altijd door een derde partij moet gebeuren. Ook is het bij het tekenen met de bestaande integraties (bijvoorbeeld Lightroom) momenteel mogelijk om met een willekeurige naam te tekenen. De keten van 'trust' is in de ogen van Demeyer dus nog niet in orde.

[82] Tom Demeyer
[83] Yivi
[84] AID Wallet
[85] European Digital Identity

- Tenslotte blijft het gissen wat de beweegredenen van de trekkers van het initiatief zijn, aangezien dit grote marktpartijen met een eigen (commercieel) belang zijn. Dit zal met het voorgaan van de tijd moeten blijken.
- De standaard is opgezet duidelijk vanuit een audiovisuele mediaketen die met hardware tot stand komt en met bewerkingssoftware richting distributie gaat, op basis van op zichzelf staande mediabestanden die via platformen worden verspreid. Hiermee lijkt het minder geschikt voor geschreven media. In het verlengde hiervan vindt hij het jammer dat HTML pagina's niet ondertekend kunnen worden met deze standaard. De standaard is alleen bedoeld voor mediabestanden, niet voor lopende tekst op een webpagina.

De belangrijkste les die Demeyer uit Proof of Provenance heeft geleerd, is zeker ook op C2PA van toepassing. Dit gaat over de consumptie van de media-uitingen. Het is namelijk een flinke uitdaging om eindgebruikers de betekenis van een dergelijk manifest met een cryptografische handtekening uit te leggen. Demeyer is van mening dat zegeltjes en 'i'-tjes voor eindgebruikers niet voldoende is. Hier ligt een ontwerpuitdaging voor de implementatie. Hier zal in zijn ogen dan ook de eerste aandacht van de NPO naar uit moeten gaan. Hij raadt aan om als eerste de discussie te hebben over wat "authentieke media" zijn? Hierbij moeten alle uiteindelijke tekenende partijen betrokken worden, zodat hier verschillende perspectieven meegenomen worden en de uitkomst van deze discussie wordt gedragen door de omroepen die hier gebruik van gaan maken. Daarbij moet deze definitie niet te breed worden.

Verder adviseert hij om in eerste instantie vooral de ontwikkelingen rond C2PA goed te volgen, maar niet te implementeren als er nog fundamentele onduidelijkheden zijn. Idealiter is er namelijk een implementatie in de gehele keten voor alle uitingen, meent Demeyer. Als een tussenfase is het ook denkbaar dat de NPO een implementatie doet die een B2B aard heeft, en het manifest nog niet actief naar consumenten communiceert, maar bijvoorbeeld tussen journalistieke (media)partijen en bijvoorbeeld burgerinitiatieven. In dat geval zijn de eindgebruikers mensen die zelf bewust op zoek gaan naar C2PA-manifesten om deze zelf te valideren/inspecteren en deze ook beter naar waarde kunnen inschatten. Zolang er echter onduidelijkheden zijn over de werking, toepassingsmogelijkheden en meerwaarde van C2PA, is het misschien wel beter om nog niets te doen. Want een halfslachtige uitrol zou het beoogde doel ook juist kunnen tegenwerken en het merk van de publieke omroep schaden.

Op het technische vlak zou Demeyer de NPO aanraden om de manifesten waar mogelijk altijd te embedden. En in de gevallen dat dit niet mogelijk is, deze op het eigen webdomein te hosten (en niet bij een cloud aanbieder). Hij is van mening dat hosting op dit herkenbare webdomein de authenticiteit verder ten goede komt, die uit de manifesten naar voren moet komen.

Tenslotte komen er nog twee losse onderwerpen aan de orde. Over de pivot van C2PA om GenAI resultaten van een manifest te voorzien, geeft hij aan dat dit prijzenswaardig is, maar ook het risico meebrengt dat wanneer zo'n manifest niet aanwezig is, men de onjuiste aanname doet dat iets dus zeker geen GenAI resultaat is. En over alternatieven op C2PA geeft hij aan dat een DNS gebaseerde aanpak in zijn ogen onveilig is. Denk aan hoe dit voor mail ook niet heeft gewerkt in de praktijk.

Dr. Neal Krawetz and Shawn Masters (English)

We interviewed Dr. Neal Krawetz and Shawn Masters as vocal critics of C2PA and promoters of the alternative specification SEAL.

SEAL (Secure Evidence Attribution Label)[86]

SEAL is an open-source specification for cryptographically signing media files. It allows web domains to sign a file, providing proof that the file was signed by an entity that controlled that web domain at a certain point in time.

SEAL is based on the principles of DKIM (DomainKeys Identified Mail), which is used to verify that an email comes from the domain it claims to be from, to ensure that the message has not changed in transit. And that the sender cannot deny that they have sent it. Like DKIM, SEAL uses a digital signature that can be validated by checking the sender's domain.

The public key of the signature is stored in a TXT field in the DNS records of a domain, just like DKIM. This means that the key is stored decentrally and the provider does not know who used the key to verify the file.

The specification supports multiple file formats, and the information can be stored differently for each file format.

SEAL and C2PA cannot (easily) be applied to the same file. A signature from one will change the file and invalidate the other.

Dr. Neal Krawetz is the founder of Hacker Factor Solutions (hackerfactor.com) where he specializes in non-traditional computer forensics, online profiling, networking, and computer security. In 2012, he started FotoForensics -- an online service for digital photo analysis. His research focuses on anti-anonymity technologies and digital abuses.

Shawn Masters is Chief Technology Officer at AELIUS Exploitation Technologies, LLC. AELIUS Exploitation Technologies specializes in analyzing big data and in advanced technology integration and insertion. AELIUS has expertise in computer security and digital forensics.

As an introduction, Krawetz explains that he came across C2PA in the context of his service FotoForensics, and was enthusiastic at first. However, after some further exploration he concluded that the specification does not live up to its claims, and introduces an undesired possibility for personal tracking instead (more on both later). Krawetz has a series of blog posts on hackfactor.com that expand on the found vulnerabilities.

Krawetz and Masters have discussed their critique of positioning of C2PA's specification and its technical vulnerabilities with Leonard Rosenthol (Senior Principal Scientist with Adobe Systems and chief architect of C2PA), who agreed with the observations, but disagreed on the conclusion regarding its utility. To Rosenthal "a broken standard is better than no standard at all".

[86] Zie ook Bijlage 4 – Secure Evidence Attribution Label (SEAL)

Some ‘bugs’ with the specification have also been addressed, in reaction to the blog posts from Krawetz. These include:

- That the ability to self-sign manifests are now marked as untrusted by the C2PA tools[87] [88]
- That the BBC has since changed the manifest for the video with the added sound[89]

Rosenthol challenged Krawetz & Masters to come up with a counter-proposal, which resulted in SEAL (also more on this later).

The positions taken regarding the C2PA specification by Krawetz and Masters are rooted in their support of the NIST definition of authenticity[90]. Authenticity requires the ability to validate. The fundamental problem with C2PA, according to them, is that C2PA's Content Credentials trusts that the metadata and corresponding content are as is claimed in the manifest. This manifest is indeed signed for validation, but the name of the signee is not controlled. Any signer can add a name to a manifest.

This means that a lot of trust is involved regarding ‘good actors’ that implement C2PA, while this same trust can easily be abused by ‘bad actors’. Krawetz argues that “it can also be exploited by bad actors.”

Masters adds that C2PA creates and promotes the assumption that the entire chain “from glass to glass” (from the glass of the lens to the glass of the content consumer's screen) can be trusted. This is a fallacy according to Masters. As an example he points at the possibility to hijack a smartphone and insert a photo as a signal, making it appear as this has been originally taken on that device. For him “authenticity is stating the trust that you have, and not implying any additional trust.” He expects that realistically the limit is proof who has signed something.

The alternative SEAL solely focuses on creating trust that you know who signed it, instead of addressing all of the problems C2PA claims to tackle. Simply put: “Publishing fake media is less likely to happen if they need to put their name to it.” This doesn’t only go for SEAL, but all digital signature systems, if implemented and positioned correctly. In summary he posits: “Glass to glass is a fallacy. It’s not going to work. Stick to proving actor X is saying Y.”

Krawetz also critiques the lack of verifiability of the Provenance part of C2PA, “provenance is where something came from”. C2PA in theory allows recording this track, but all of the previous signatures can be changed when a new C2PA-manifest is signed. So again trust in the good faith of the final actor is implied. In short: “The user base needs to understand what a signatory system does. C2PA overstates what it does.”

To Krawetz and Masters, a strength of SEAL vs. a weakness of C2PA lies in its respective simplicity vs. complexity. SEAL doesn’t claim trust beyond the owner of a domain claiming the content. It uses proven distributed technology (DNS & public-key cryptography), that if

[87] Content Credentials : C2PA Technical Specification: The Authority Key Identifier extension shall be present in any certificate that is not self-signed, as per RFC 5280, section 4.2.1.1.
[88] See update on C2PA's Worst Case Scenario - The Hacker Factor Blog
[89] Haiti violence: Haiti gangs demand PM resign after mass jailbreak
[90] authenticity - Glossary | CSRC

supported can function as a notary to validate who is responsible for a signed file (the one with control over the domain).

SEAL is not as feature rich as C2PA claims to be. It only allows for hard bindings, which further hardens the claim. Additionally, SEAL does not address the problem of metadatastripping. They see this as a strength, a valid signed file has not been changed by a third party.

Masters sees a SEAL vulnerability in "typo domains" to mislead validators. However there is still a trail of ownership from that typo domain. Another vulnerability is "domain hijacking", when someone takes control over a domain without the owner realising. However with SEAL, you also need a private key that only the signer should have. This means that the party who has taken control needs to add its own public key, making the claims that were done in bad faith recognisable. Additionally previous signatures can be revoked after repurposing the domain. Finally, when asked if C2PA and SEAL could work in tandem, Krawetz responds by pointing out that the solutions are technically incompatible, as they both need to modify the file.

Returning to C2PA: Krawetz and Masters say that it can still be fixed, but only if complexity is reduced instead of expanded. They add that the current implementation is monolithic, lacks transparency (especially the Web Assembly code for the validation) and involves an expansive Software Bill of Materials, which introduces a lot of dependencies on external code (and possible vulnerabilities). They also fear the specification may expose its adopters to patent issues down the line.

They go on to explain that while C2PA is open in theory, given the issues mentioned above, and the complexity in compiling and adopting the open source components yourself, they expect that in practice most users will in fact adopt the proprietary tools offered by parties like Adobe. They see an analogy with how the PDF standard was introduced by Adobe and allows other PDF viewers, but in practice most people use the Adobe Acrobat products to view/edit PDF files.

They point out that this also introduces a privacy issue. Again using the analogy of the PDF standard, they point out that Adobe Acrobat 'calls home' to Adobe for a lot of user interaction with PDF files in their proprietary software. This is even the case for the PDF/A archival file format (later versions of the ISO standard). Adobe and third party providers has made a substantial business of selling tracking of documents for review and process control which relies on information of when documents are viewed to be collected by central servers outside of the users control.[91] They argue that the freely-offered Adobe tools to sign and verify C2PA-manifests goes down the same path and that users need to be aware of potential data collection around the automated verification in C2PA. Since the C2PA specification is being fast tracked to become an ISO standard, this risk of tracking and privacy issues is increased.[92] Rosenthol, chief architect of C2PA, is reviewing C2PA for becoming an ISO standard. Masters and Krawetz point out the apparent bias and conflict of interest in having C2PA chair the ISO review of C2PA's specifications.

In addition, they point out that more recently they've seen C2PA pivot from positioning the proposed as a way for authors to sign their expressions, to a means for GenAI providers to mark

[91] Adobe Stock Breakdown: How Does Adobe Make Money In 2022?
[92] See #c2pa #contentcredentials | Leonard Rosenthol | 14 comments

their artificially generated media as such. They feel this is done to show self regulation by GenAI parties, but also to address issues with entropy and eventual model collapse (AI models being trained on AI generated data).

When asked about their experience with the community around C2PA, they mention that there is an active developer community (through a Discord channel), where a lot of the issues pointed out in the blog posts by Krawetz are openly discussed and widely known. However, it is also their experience that these issues are being downplayed by the Adobe developers. As the user end of the community, they feel that a lot of participants are simply "drinking the Kool-Aid", and in some cases have already made business decisions they can't easily undo.

Turning back to the challenge of public broadcasting, authenticity of media expressions, and battling disinformation, Krawetz and Masters see the following path forward: "Broadcasters should pick a system to sign all their expressions. Let this be an archival record that can be verified." In addition they should make a best effort at their end to "verify the sources that they use on the ingest".

They illustrate this example by imagining that a programme like the PBS News Hour should only use local news reports from 'stringers' that can be verified, and on distribution publish their verified sources and the (signed) final product, so this can be checked by interested individuals. In this context they mention the valuable work being done by fact-checking initiatives like Snopes and AFP. Adding a layer of signatures that can validate who has published something has the potential to "democratise fact-checking". This is in contrast to a media environment where people share media that they've "found on YouTube (so it must be true)". Masters mentions in this context that "conspiracy theories are the threat to modern society".

Prof. dr. Yael de Haan

"[Prof. dr.] de Haan is lector Kwaliteitsjournalistiek in Digitale Transitie aan de Hogeschool Utrecht en bijzonder hoogleraar Lokale Publieke Omroep aan de Rijksuniversiteit van Groningen. Ze leidt een onderzoeksgroep die onderzoek doet naar welke rol journalistiek speelt in de levens van mensen en de maatschappij. Inzicht in het publiek en de samenleving biedt handelingsperspectieven voor de journalistieke beroepsgroep. Technologische ontwikkelingen zijn hier onlosmakelijk mee verbonden. Het onderliggende vraagstuk draait om vertrouwen: vertrouwen in de journalistiek, in elkaar, in de samenleving en in technologie. Onderzoeksprojecten gaan in op de rol van AI in de journalistiek, de veranderende nieuwsconsumptie van diverse publieksgroepen en de rol van ethiek in een veranderende journalistiek.[93]

Ter introductie beschrijft De Haan hoe zij een spanningsveld in de stand van kwaliteitsjournalistiek constateert. Sinds de jaren '80 is er een professionalisering van het open beroep journalistiek op gang gekomen, deze heeft geleid tot een standaard en instrumentarium dat gericht is om betrouwbaar en geloofwaardig te werken. Denk daarbij aan zaken als hoor en wederhoor, gebruik van meerdere bronnen, een journalistieke code, verschillende opleidingen en de Raad van Journalistiek. Deze professionalisering heeft de beroepsgroep verder ontwikkeld, maar tegelijkertijd kan het ook voor een afstand tussen de journalist en de samenleving zorgen.

Het publiek heeft andere verwachtingen van kwaliteitsjournalistiek, dan de journalisten zelf hebben.[94] De verwachtingen van het publiek rondom representativiteit en vertegenwoordiging van perspectieven zijn veel breder. Deze beweging komt met name voort uit de hoeveelheid informatie die voor burgers beschikbaar is. Wanneer een burger zijn perspectief mist bij een media-uiting kan dit leiden tot vermindering in vertrouwen. Hierdoor werkt een "keurmerk" voor kwaliteitsjournalistiek niet meer. De afgelopen vijf jaar is dit besef ook binnen de journalistiek gegroeid. Dit speelt ook in andere sectoren, zoals bij artsen die door hun patiënten met eigen onderzoek op het internet en daaruit opgedane inzichten en conclusies worden geconfronteerd.

De Haan ziet de veranderende verwachtingen van de burger op dit vlak als een verworvenheid van de digitale transitie: "De burger is slimmer dan ooit. Er is zoveel informatie ter beschikking." Ook wijst ze er op dat, hoewel er terecht zorgen zijn over 'filterbubbels', er geen hard bewijs is dat mensen onvoldoende informatiebronnen tot zich nemen. Ook uit longitudinaal onderzoek naar nieuws-mijders[95] dat De Haan aan het uitvoeren is blijkt dat dit desondanks hun bewuste keuze om geen 'nieuws' te consumeren, dit toch geïnformeerde burgers zijn.

De Haan wijst op een aantal nieuwe uitdagingen van deze tijd: fake news, eenzijdige berichtgeving en burgers die niet gehoord worden. Zij stelt daarbij: "Over het algemeen vinden mensen het moeilijk de complexiteit te begrijpen die achter berichtgeving schuilt." Kan je bijvoorbeeld het recente gesprek (maart 2024) in het Oval Office tussen Trump en Zelensky reduceren tot een onbeschofte opstelling van Trump ten aanzien van een leider van een land in nood? De uitdaging is waar de burger dan vervolgens op zoek gaat naar de waarheid? Maar,

[93] https://www.hu.nl/onderzoek/onderzoekers/de Haan-de-haan
[94] Zie Journalismlab – Nieuwsmijding en het effect op de burger en nieuwsmedia in een democratische samenleving
[95] Artikel binnenkort beschikbaar

benadrukt De Haan: "Waarheid is sowieso een lastig begrip, maar het begint met het aan bod laten komen van de verschillende perspectieven."

Kijkend naar vertrouwen in de journalistiek, zoals bijvoorbeeld blijkt uit de jaarlijkse rapporten van Reuters[96], geeft De Haan aan dat de mate van vertrouwen in journalistiek bij de burger uitvragen problematisch is. De bevraagden kunnen verschillende definities van journalistiek hanteren. Bij vertrouwen in journalistiek zijn twee factoren van belang, is haar ervaring: "Erkenning en herkenning". Met andere woorden: Mensen moeten de berichtgeving kunnen volgen en kunnen inzien wat de relevantie ervan is voor hen. Kijkend naar het eerdergenoemde onderzoek naar nieuwsmijders, blijkt overigens dat maar een kleine groep (8%? van nieuws-mijders[97]) het nieuws mijdt uit wantrouwen.

De consument van journalistieke uitingen ervaart vooral een mismatch tussen de verhalen over de maatschappij en hun eigen belevingswereld. Ook dit is een breder probleem, wat bijvoorbeeld ook in de politiek terug te zien is. De Haan is ervan overtuigd dat hier een kans voor de journalistiek ligt om meer de maatschappij in te trekken.

Tegelijkertijd zijn burgers op sommige onderwerpen (noodgedwongen) experts geworden. Denk bijvoorbeeld aan de gaswinning in Groningen. Een journalist van een landelijk medium die "een dagje komt kijken" verliest het dan van de expertise van burgers te plekke. Dit roept de vraag op waar de autoriteit ligt? De uitdaging voor de journalistiek bestaat dan ook uit het meenemen van (de perspectieven van) deze 'externe' experts. Lokale media zijn wat dat betreft iets beter gesitueerd, maar zijn tegelijkertijd minder uitgerust om diepgravend onderzoek te doen en kunnen zich ook nog beter inbedden.

Gevraagd naar haar definitie van authenticiteit (in de context van journalistieke uitingen) geeft De Haan allereerst aan dat ze de term weinig gebruikt. Voor haar zijn authentieke journalistieke uitingen afkomstig van de journalist zelf (zelf vervaardigd). Uitingen die zijn geaggregeerd zijn niet authentiek en auteursrechtelijk complex.[98] Maar tegelijkertijd vraagt ze zich af of er nog sprake is van authenticiteit "wanneer iedereen dezelfde invalshoek pakt".

Verder wijst ze op het belang van de identiteit van de afzender. Deze is gelaagd. Van de journalistieke/media sector, naar het publieke bestel, naar specifieke omroepen en tot aan de individuele journalist. Deze laatste is tegenwoordig vaak ook een "personal brand", dat via sociale media wordt uitgedragen. Deze identiteiten kunnen botsen. Je ziet dat media hiermee spelen. Denk aan hoe Follow the Money en de Correspondent de identiteit van journalisten inzetten als afzender van bepaalde uitingen (nieuwsbrieven, boeken, podcasts, etc.). Tenslotte noemt De Haan dat er uit onderzoek overigens blijkt dat de afzender van een uiting niet altijd een rol speelt in het oordeel van de consument.

Hoewel De Haan een verband tussen authenticiteit en vertrouwen ziet, is deze er niet met "waarheid" (waarvan ze eerder al aangaf dat dit sowieso een problematisch begrip is). Het is

[96] Digital News Report Nederland 2024: Interesse in nieuws neemt af, vertrouwen in nieuws daalt licht - Commissariaat voor de Media
[97] De Bruin, K., Vliegenthart, R., Kruikemeier, S & de Haan, Y. (2024): Who Are They? Different Types of News Avoiders Based on Motives, Values and Personality Traits, Journalism Studies, DOI: 10.1080/1461670X.2024.2321537
[98] Journalismlab – De 50-woordenwet: aggregatie voor dummies

mogelijk dat een authentieke uiting tegelijkertijd een 'onwaarheid' is. Ze wijst erop dat de gemiddelde consument ook begrijpt dat absolute objectiviteit een fictie is. Het is de claim van de journalistiek op deze objectiviteit die voor weerstand zorgt. Er is een behoefte om te begrijpen hoe een uiting tot stand is gekomen. Denk aan een podcast, waarin een journalist de luisteraar meeneemt in de totstandkoming van het verhaal.

Volgens De Haan is de adoptie van nieuwe technologieën (zoals digitale handtekeningen) door journalistieke redacties essentieel voor de slagingskans. Anders gebruiken alleen de medewerkers met affectie voor "technologische gadget" het instrument. Er moet voor deze redactie een duidelijk probleem (van journalistieke aard) worden opgelost. Daarnaast wijst De Haan erop dat de werkprocessen, vanwege de professionalisering vanaf de jaren '80, strak zijn ingericht en weinig ruimte bieden voor aanpassing. Dit wordt vaak gezien als "extra werk". Om het te laten slagen moet het een logisch onderdeel van de bestaande rituelen zijn en uiteraard ook gebruiksvriendelijk.

Meer specifiek kijkend naar de belofte van C2PA, heeft De Haan twijfels bij de meerwaarde voor de burger of consument. In haar ervaring is de kwaliteit van de uiting (eenzijdig, gestuurd) meer een reden voor wantrouwen, dan het feit dat de afzender niet geverifieerd kan worden door de lezer.

Voor redacties noemt ze een middel als C2PA vooral een "zelfreflectief instrument". Luisterend naar de omschrijving van interviewers ziet ze geen onafhankelijke partij die de kwaliteit van de claims controleert. Maar het vastleggen en tekenen voor de herkomst en authenticiteit van een uiting dwingt de journalist wel om in verschillende stappen in het proces bewuste checks toe te passen en vast te leggen. Het zijn dan wel de managers die dit moeten afdwingen als werkwijze.

Deze zelfreflectie voor de journalist zou vervolgens richting de consument als een instrument voor transparantie over de totstandkoming van de uiting ingezet kunnen worden. Het publiek is volgens De Haan geïnteresseerd in een (gesimplificeerde) ontleding van de stappen waarmee een journalistieke uiting tot stand komt, en welke checks hierbij om de hoek komen kijken. Dit zou een vorm van PR voor de journalistiek kunnen zijn om meer vertrouwen te scheppen. Hier ziet De Haan veel meer in, dan in een vinkje om de authenticiteit te bewijzen.

Wanneer tenslotte gevraagd wordt naar de rol van de initiatiefnemers van de specificatie (waaronder een aantal grote technologische spelers) en de houding van journalisten ten opzichte van dit soort partijen, vraagt zij zich vooral af wat de implicaties van de verwerking van de data zijn? Waar wordt alle data opgeslagen? Is dat wenselijk vanuit journalisten gezien? Daar zullen naar haar verwachting zeker vragen over zijn onder deze doelgroep.

Code Journalistiek Handelen voor de Landelijke Publieke Omroepen
Zelf schreef De Haan mee aan de recente (2023) Code Journalistiek Handelen voor de Landelijke Publieke Omroepen[99], ook een voorbeeld van het instrumentarium dat is ontwikkeld sinds de jaren '80. De Haan benadrukt dat deze code niet spreekt van "objectiviteit" en dat de code gericht is op medewerkers van de publieke omroepen, niet alleen journalisten. Zo vraagt een uiting als een talkshow ook om journalistiek handelen.

[99] Code Journalistiek Handelen

Prof. dr. ir. Jeroen de Ridder

Prof. dr. ir. Jeroen de Ridder is hoogleraar politieke epistemologie aan de Vrije Universiteit. Samen met prof. dr. Rik Peels schreef de Ridder het recent uitgebracht boek "Wat is nou waar?".

Als achtergrond voor zijn boek geeft de Ridder aan dat sinds 2016 het post-truth denken de wereld bezig houdt. Wie kan je vertrouwen? Desinformatie over COVID-19 en generatieve AI worden gezien als bedreigingen die de informatieomgeving "vervuilen". Daarnaast zijn er verschillende psychologieboeken uitgebracht die aangeven dat we "slecht kunnen denken". Mensen zijn biassed en beperkt. Dat is in zijn ogen een te sombere visie.

Hij vult aan dat de stelling dat we omkomen in desinformatie overdreven is. Maar door de beschikbaarheid en veelheid van informatie, is desinformatie wel makkelijker te vinden. Met name doordat mensen (digitale) gemeenschappen vinden, waarin eigen denkbeelden versterkt worden. Vroeger stonden deze mensen met 'afwijkende' denkbeelden meer alleen en vonden ze meer tegenspraak (in de kroeg of het koffiehuis).

Volgens de Ridder klopt het niet dat mensen niet kunnen denken. de Ridder en Peels willen een constructieve benadering en schreven daarom een "epistemologisch zelfhulpboek". Zonder overigens het woord epistemologie te noemen. Ze kijken naar de verworvenheden op het gebied van kritisch denken over de eeuwen heen: de filosofie, de journalistiek, de wetenschap, etc. "Die wijsheid der eeuwen kan je toepassen in je eigen leven."

Na een introductie over C2PA gaat het interview over op de term authenticiteit. Naast epistemologie wordt ook authenticiteit niet genoemd in het boek. de Ridder denkt bij dit begrip bijvoorbeeld aan bronnenonderzoek als methodiek: Je afvragen wie de bron van informatie is en wat de consequenties zijn van deze bron. Welke belangen heeft de bron, welke beperkingen kennen ze?

In het onderzoek naar C2PA komen begrippen als authenticiteit, waarheid, vertrouwen, en autoriteit veel naast elkaar voor. Samen met de Ridder verkennen we deze begrippen, ook in relatie tot elkaar.

Authenticiteit volgens de Ridder is niet gekoppeld aan waarheid. Authenticiteit heeft voor hem te maken met een antwoord op de vraag waar een uiting vandaan komt, of een persoon of organisatie echt de bron is van die uiting. Dat staat los van de vraag of zo'n bron ook de waarheid spreekt of betrouwbaar is. Met een bewijs van authenticiteit kan de consument dan verschillende kanten op gaan: "Daar komt het vandaan, dan negeer ik het," als zij/hij de bron al niet vertrouwt. Of andersom: "Ik consumeer bewust alleen berichten van die bron omdat ik die betrouwbaar acht". Authenticiteit bevestigen van een bron kan de drempel verlagen om de bron te verifiëren en deze kwalitatieve vragen te stellen. Omdat vertrouwen in de gevestigde media is gedaald, zou verificatie van herkomst dit vertrouwen weer (deels) kunnen terugwinnen.

de Ridder ziet verificatie van authenticiteit als een stap in de principes van transparantie en controleerbaarheid (van journalistiek). Hierbij verwijst hij ter vergelijking naar open science. Er is wel een substantieel verschil in hoe de journalistiek hiermee omgaat en hoe de wetenschap daarmee omgaat. In de wetenschap wordt voor tamelijk radicale openheid gepleit, anderen kunnen vanaf de bron meekijken in het wetenschappelijk proces en het zo in principe zelf

controleren. Maar in de journalistiek zie je dit soort transparantie niet in dezelfde mate. Bijvoorbeeld bij het publiceren van ruwe data (en er zijn vaak ook goede redenen voor om zulke data in de journalistiek niet zomaar te delen, bijvoorbeeld bescherming van bronnen en privacy). Vertrouwen in wetenschap scoort gemiddeld wel wat hoger dan journalistiek[100], hoewel de Ridder twijfelt of dit door open science principes komt. Daarvoor is de open science beweging nog te jong en kleinschalig.

de Ridder reflecteert vervolgens over de relatie tussen authenticiteit en autoriteit. Hij verstaat onder autoriteit een entiteit die gezaghebbend is in het maken van claims. De NOS kan bijvoorbeeld als een autoriteit gezien worden in de context van een bepaald type (nieuws)uiting. Economie, binnenlandse politiek, voetbal. Waar de gebruiker kan aannemen dat berichtgeving volledig, gecheckt, en genuanceerd is (waarheid in brede zin). Een entiteit kan een autoriteit op het ene zijn, maar dat betekent niet dat deze overal een autoriteit op is.

Kijk naar de techniek (van C2PA) en authenticiteit. De interviewers geven aan dat technologische authenticiteit bevestigd moet worden door een autoriteit. In hoeverre is dat mogelijk? de Ridder geeft aan dat we dit soort vertrouwen in een autoriteit ten grondslag ligt aan heel veel dingen in de samenleving. Denk bijvoorbeeld aan bankbiljetten, deze dragen technische middelen of kenmerken met zich mee om te bevestigen dat iets geld is, maar ook dan is er een autoriteit die aangeeft hoe geld eruit ziet. En er is vertrouwen in die autoriteit.

In een ideale wereld zou de technische autoriteit (in het geval van een mechanisme om de authenticiteit van een media-uiting te verifiëren) bij onafhankelijke partijen liggen. Hierdoor zijn de technieken controleerbaar en is er geen één autoriteit die hierin kan morrelen. Dit past in een klassieke en bredere discussie over het eigenaarschap van de informatieruimte. Dat C2PA rust op de ondersteuning van big tech helpt niet voor het vertrouwen in de authenticatietechnieken. Je wil garanties voor de afscherming van de data en de onmogelijkheid van manipulatie. Hiervoor is transparantie nodig.

de Ridder herhaalt dat authenticiteitsmechanismes geen "geen shortcut naar de waarheid" zijn. Authenticiteit bevestigen is (slechts) de eerste stap. Dan moet je jezelf als consument vervolgens afvragen of je deze afzender vertrouwt. Maar de Ridder vraagt zich ook af wat je op dit vlak redelijkerwijs kan verwachten.

In een reflectie op huidige technologische ontwikkelingen beschrijft de Ridder dat videomateriaal vroeger als relatief 'keihard' bewijs werd gezien, wel in aanmerking nemend dat er altijd mogelijkheden waren om manipulatief te monteren. Dat gaat met de opkomst van generatieve AI en videosynthese niet meer op, waardoor bevestiging van de bron belangrijker wordt: "Om het in de echte werkelijkheid te verankeren, in een tijd waarin het makkelijker is geworden om nepberichten te produceren."

Breder kijkend geeft de Ridder aan dat omroepen aan eigen standaarden moeten voldoen, een goede reputatie moeten hebben, en dus ook actief met reputatieschade om moeten gaan.

Wanneer gevraagd over andere methodes om desinformatie tegen te gaan, waaronder factchecken, geeft de Ridder aan dat hij hier inmiddels in toenemende mate sceptisch over is. Hoe goedbedoeld ook, het bereik en effect van factchecken blijkt beperkt in de praktijk. Ook over

[100] zie bijvoorbeeld: Vertrouwen in de wetenschap | Rathenau Instituut

‘pre-bunken’ is de Ridder sceptisch. Het kan inderdaad helpen om een kritischere houding te ontwikkelen, maar onderzoek laat zien dat deze houding gemakkelijk doorschiet en een impact heeft op de beleving van betrouwbare informatie: mensen gaan niet alleen mis- en desinformatie minder geloven, maar ook goede, betrouwbare informatie. Het kan daardoor mogelijk “kleine complotdenkers” van mensen maken.

De uitdaging is, volgens de Ridder, daarentegen meer om mensen te laten geloven in betrouwbare informatie. Ook gezien de daadwerkelijke (nog best beperkte) schaal en impact van desinformatie. Daar kan openheid, transparantie, en uitleg van processen bij helpen. Realistisch gezien ligt het laaghangende fruit bij de uitleg over welke principes nu worden gehanteerd en waar men met klachten terecht kan. Voor vertrouwen is de intermenselijke relatie in de praktijk ook van belang: “Wat je kent, vertrouw je meer.”

Het hele journalistieke proces vastleggen, zoals bij open science, is waarschijnlijk te tijdrovend binnen de journalistieke praktijk. De wetenschap kan hier tijd voor nemen, de journalistiek niet. We kijken kort naar Wikipedia als voorbeeld wel een optimum van verifieerbaarheid, maar dat is makkelijker in die specifieke context.

Lucille Verbaere and Mohamed Badr Taddist (EBU) (English)

Lucille Verbaere

Lucille is Senior Project Manager at the European Broadcast Union, Technology and Innovation, coordinating Media Data Space and Cybersecurity activities. Verbaere is part of Project Strategic Board of Tems, for the deployment of a Trusted European Media Data Space[101], Digital Europe programme). She has 20y-experience in Cybersecurity, Telecommunications, Air Transport and Semiconductor industries. She started as an R&D engineer doing research on wireless telecom systems, then moved to program and product management positions for connectivity and data management services to airports and airlines. Before joining EBU in September 2020, Verbaere was responsible for a portfolio of cybersecurity products, based on quantum physics.[102]

Mohamed Badr Taddist

Mohamed Badr Taddist is a Cybersecurity student at EPFL-ETHZ and an intern at the EBU. He works on Leveraging C2PA to combat misinformation and authenticate content provenance for news publishers and broadcasters' workflows.

Summary

Verbaere introduces that she works at EBU on the topics cybersecurity, provenance & data spaces. This currently includes many C2PA integration projects. Taddist is an intern at Technology & Innovation at EBU and has been working on different provenance technologies and communities, including C2PA, but also JPEG Trust.

When asked for some background on the involvement in C2PA, Verbaere starts off by mentioning that the first use case for C2PA is to flag AI generated content. She goes on to say that EBU members are all concerned with misinformation, and the role GenAI plays to foster this. “Members want to use trusted information in their news rooms”. In this B2B scenario the members want to identify “real information” from the source. As we now live in a media landscape where “fake news is easy to generate”.

She then explains that there is also a second use case for their members. In this B2C scenario members want to brand their expressions with a “trusted stamp”. Both are important to members. But in Verbaere’s experience they currently focus mostly on this B2C, but also acknowledge that user adoption is key here. The EBU members are in a phase of innovation and experimentation with implementing C2PA currently.

When asked for a definition of “authentic media expressions”, Taddist states that “every expression that can be verified and its source is trusted is authentic”. He adds that these statements are not necessarily relying on the truth. With C2PA, consumers can verify and trust the expression is received as the publishing actor intended.

When asked about their experience with the EBU workflow demonstrator, Taddist states that this has shown that the C2PA specification works well and is interoperable. It has demonstrated that

[101] Tems
[102] Lucille Verbaere PublicSpaces Conferentie

it is practically possible to implement every step in the workflow (from glass to glass), when using supported hardware and software. He then stresses that in reality it "will take time". This requires adoption of the specification by many actors, who are all involved, to have an entire workflow (for public broadcasters) covered.

The demo was currently limited to using a Leica (still image) camera and the Adobe media editing suite of software. Since then a few more camera brands have started adopting C2PA. Verbaere rephrases this, by stating that working with C2PA currently requires "pragmatism", when implementing the specification. For her it is also still to be decided which information about the provenance is relevant for different stakeholders for ex journalists in news rooms versus public.

When asked about the potential data privacy issues related to the use of validators like ContentCredentials.org and how current implementations deal with trusted Certificate Authorities, Verbaere underlines that in her view "the current specification is still open" and that everything is still "in progress". She states that the EBU tries to exercise influence to ensure the interests of public service media (and their users) are being covered by C2PA. She is also aware of the steps that are being taken to turn the specification into an ISO standard (JPEG Trust Part 1 which is known as core foundation (ISO/IEC 21617-1:2025) is already published as a standard under ISO and IEC, and builds upon C2PA).

Looking at potential data privacy issues from a technical perspective, Taddist shows that ContentCredentials.org utilizes a list of trusted certificates[103] and that a certificate that is not on that list is shown as 'unknown' when validated by the website. This includes self-signed certificates. Other verification tools have different trust lists and policies (like IPTC Origin Verified News Publishers List., or Origin Verify[104]). [note: Origin Verify uses an additional trust list next to the list of ContentCredentials.org. This list, IPTC Origin Verified News Publishers List, adds certificates from six news agencies].

Taddist confirms that the verifiers check the validity of the certificate, as part of the display and validation of the C2PA-manifest. As any web-based application, a verifying application could lead to the disclosure of information, including personal information on the person who wants to verify content, the content itself and its descriptive information. This can be used to create a profile on the person verifying content. Taddist confirms that other attack vectors on C2PA are possible. E.g. people could be misled by a bad faith verifier. Users can ignore signals of verification or trust the direct presentation of results of material with a CR-mark. Taddist recognises that "technology education" for users is needed. As is governance, when it becomes evident that tampering is happening. He concludes: "The technology is safe, but there are some tricks". Taddist clarifies that safe should be read within the defined scope of C2PA, "in the sense that it can detect manifest and content tampering. For example, I cannot change the claims that have been inserted in a C2PA-manifest created by the BBC or impersonate the BBC itself". Taddist further points to potential harms that have been recognized by C2PA itself.[105]

[103] https://contentcredentials.org/trust/allowed.sha256.txt
[104] IPTC Origin Verified News Publishers List
[105] C2PA Harms Modelling

Taddist also points to some limitations of the C2PA specification, related to the use cases that are relevant for broadcasters. Currently “contextual or journalistic information” about the content (like fact-checks, etc.) cannot be displayed, only edits and corresponding authors. Taddist also remarks that there is a drawback to how redactions are dealt with in the C2PA spec. Hybrid Blockchain could be introduced here to deliver transparency and traceability for the provenance data, but with the downside that the information shared then would be immutable. As a journalist’ safety may be at risk, such Distributed Ledger Technology should be governed by a neutral, trustworthy entity that can strongly safeguard democratic values.

Verbaere returns to the purpose of using C2PA: verifying from where something came from, it requires you to trust that source. She sees the opportunity for broadcasters to “brand themselves as trusted”. However this also puts more burden on the signer. Proving the authenticity means that your metadata is authentic, errors in the provenance information creates a risk that you lose credibility.

C2PA in de EBU News Pilot

De European Broadcasting Union (EBU) heeft in een project (EBU News Pilot) C2PA geïmplementeerd om de transparantie van beelden op het platform te vergroten. Nieuwe beelden op de News Pilot krijgen een elektronisch ondertekend C2PA-manifest. Hiervoor gebruiken ze de CAI-tools. Het project moedigde leden aan om verrijkte herkomstinformatie aan te leveren. Het project kijkt naar de ondersteuning van C2PA in audio en video.[106]
De EBU onderzoekt samen met leden hoe C2PA-informatie aan eindgebruikers getoond kan worden, bijvoorbeeld in het kader van "A European Perspective," een project voor de uitwisseling van digitaal nieuws.

[106] Zie EBU joins early implementers of new standard to track media content provenance

Bijlage 4 – Secure Evidence Attribution Label (SEAL)

Het PoP verslag spreekt van enkele alternatieven, in deze evaluatie van C2PA zijn ze minder relevant, De alternatieven van PoP zijn technisch veelal niet equivalent aan C2PA. In deze bijlage wordt het meest relevante alternatief besproken. Secure Evidence Attribution Label (SEAL).[107]

SEAL is ontwikkeld door Dr. Neal Krawetz (Hacker Factor). De specificatie is een open source-oplossing voor het cryptografisch ondertekenen van mediabestanden. Hiermee kunnen entiteiten een bestand ondertekenen. Deze handtekening bewijst dat het bestand op een zeker moment door de entiteit ondertekend is.

De specificatie is gebaseerd op de principes van het veelgebruikte DKIM (gebruikt voor e-mail). DKIM is een manier om te verifiëren dat een e-mail afkomstig is van het domein dat het beweert te zijn. Het gebruikt een digitale handtekening die kan worden gevalideerd door het domein van de afzender te raadplegen. Dit helpt om e-mailspoofing en phishingaanvallen te voorkomen.

De publieke sleutel van de handtekening wordt net zoals bij DKIM in de DNS van een domein opgeslagen. Hierdoor wordt de sleutel meteen decentraal opgeslagen en weet de aanbieder niet wie de sleutel gebruikt heeft om het bestand te verifiëren.

De specificatie ondersteunt meerdere bestandsformaten, per bestandsformaat kan de informatie anders opgeslagen worden.

Net als C2PA bevat een handtekening meerdere velden, waaronder:

Technische velden – technische velden als de gebruikte versleuteltechniek, versie, gebruikt hashing algoritme, etc. helpen het bestand uniek te identificeren. Het veld domein ‘d’ geeft het domein, bijvoorbeeld ‘npo.nl’ aan waar de in de DNS de public key opgeslagen is.

Metadata – velden als ‘uid’ ‘id’, ‘copyright’, ‘info’ geven de ondertekenaar de mogelijkheid om optioneel additioneel informatie toe te voegen aan de handtekening.

Handtekening – het veld ‘s’ slaat de handtekening op die met de public key uit de DNS van het domein ‘d’ geverifieerd kan worden.

Uit een beperkte SWOT-analyse van SEAL blijkt dat deze veel gelijk heeft met de SWOT van PoP. Het deelt zowel de kracht als de zwaktes van de specificatie.

Strengths – SEAL heeft een sterke technische achtergrond, heeft weinig technische infrastructuur nodig en maakt gebruik van bestaande breed geadopteerde principes.

Weaknesses – SEAL lijkt een theoretisch antwoord op C2PA, heeft (nog) geen brede adoptie of community. Daarnaast heeft het nog geen gebruikersinterfaces en door de hard bindings kan het een handtekening snel verloren gaan (op social media bijvoorbeeld). Ook hier is metadata alleen beschikbaar binnen een bestand.

Threats – Geen adoptie betekent dat verificatie en gebruik niet hoog zal zijn.

[107] Zie hackerfactor/SEAL: Secure Evidence Attribution Label (SEAL)

Bijlage 5 - Beschikbare tools en implementaties

Deze bijlage is een kopie van de implementaties die gevonden zijn door Badr Taddist (EBU).

Tools	**Scope**	**License**	**Format supported**	**Links**
CAI SDK	Creating & signing manifests. (C2PA v2)	Open source	Images, Audio, Static Video, ISOBMFF	• CAI open source SDK • Content Authenticity Initiative · GitHub
CAI Verify	Verify content credentials, display Manifests	Free tool	Image, Audio, static video	• Verify Content Credentials
Truepic Validate & Display	Verify and displaying content credentials	Free library	Static video, and images	• https://lens.truepic.dev/docs/display-library
DASH.js	Display and verify	Open source	DASH	• contentauth/dash.js
Truepic Signer[108]	Capturing, creating & signing manifests (C2PAv2?) Certificate authority	Proprietary	Image, Audio, static and progressive video (next update)	• Truepic
IPTC Origin Verifier	Verify and display Content credentials	Free tool	HEIC, HEIF, JPEG, PNG, SVG, TIFF, WebP, MOV, MP4	• Origin Verify
Microsoft Content Integrity	Verify and display Content credentials	Free tool	Image, Audio, static video	• Microsoft Content Integrity
Web-browser display plugins	Verify and display	Free/open tools	Images, video, audio	• Browser extension to validate C2PA digital assets • GitHub - digimarc-corp/c2pa-content-credentials-extension • ContentLens C2PA Validator - Chrome Web Store
TrustNXT–c2pa	Sign and verify manifests	Open-source	Images,	• @trustnxt/c2pa-ts – npm • TrustNXT/c2pa-ts

[108] Mohamed Badr Taddist (EBU) merkt op (per mail) dat Truepic deze dienst heeft opgeheven, maar dat de Certificate Authority die voor eerder via deze dienst uitgebrachte certificaten nog wel werkt.