C2PA :Garantir l'Authenticité des Contenus Numériques

Sources :
Documentation officielle :
https://c2pa.org/
https://c2pa.org/specifications/specifications/1.0/specs/C2PA_Specification.html
Articles :
https://www.linuxfoundation.org/blog/how-c2pa-helps-combat-misleading-information
Outils :
https://contentauthenticity.org/

Semestre 5 - présenté par Hugo Fairon

INTRODUCTION AU C2PA

Le **C2PA** (Coalition for Content Provenance and Authenticity) est une initiative visant à garantir **l'authenticité** des contenus numériques . En intégrant des métadonnées, le C2PA permet aux utilisateurs de vérifier **l'origine et l'intégrité** des fichiers numériques, renforçant ainsi la confiance dans les médias et au près du publiques.

C2PA est un projet de Joint Development Foundation, formé grâce à une alliance entre Adobe, Arm, Intel, Microsoft et Truepic.

2021

- **Création de la C2PA** : Collaboration entre Adobe, Microsoft, BBC, et Truepic.

Janvier 2022

- **Première version des spécifications** : Introduction d'un standard pour la provenance des contenus.
- **Nouveaux membres** : Intel, Arm, et d'autres partenaires stratégiques.

2022-2023

- **Adoption élargie :** Sony, Twitter, Nvidia rejoignent l'effort.
- **Diversification des formats :** Ajout de nouveaux formats multimédias (audio, vidéo, texte).

2023-2024

- **Nouvelles versions des spécifications** : Amélioration de la norme avec plus de précision et d'outils de vérification.
- **Expansion des partenariats** : Getty Images et d'autres acteurs des médias.

Pourquoi le C2PA ?

Avec la montée de la **désinformation** et des **deepfakes,** il est crucial de disposer de mécanismes pour assurer **l'authenticité** des contenus.

Le C2PA répond à ce besoin en fournissant un cadre qui aide à identifier les sources et à prévenir la **manipulation et la désinformation**.

- Explosion des contenus numériques sur internet
- Prolifération des fausses informations
- Arrivée des IA (détourner de leur utilisation de base)
- Manipulation croissante des images et vidéos
- Besoin d'une norme de confiance et de transparence

Le renforcement en cybersécurité à pousser les personnes mal intentionnées à exploiter la faille la plus vulnérable. C’est-à-dire nous, les utilisateurs en les manipulant via le contenu qu’ils consomment.

Quel est la solution ?

C2PA est un standard technique que toute plateforme, entreprise ou créateur peut utiliser pour intégrer des données de provenance dans les contenus numériques. Elle agit comme un **organe de standardisation** qui offre une feuille de route technique pour la traçabilité et la transparence des contenus.

ContentAuthenticity.org (CAI) quant à lui, se concentre sur l'**implémentation** des technologies de provenance dans des **produits spécifiques** (principalement ceux d'Adobe) pour s'assurer que les contenus créés avec ces outils peuvent être authentifiés et traçables. (outil largement utilisé pour vérifier le contenu)

Actuellement seul les entreprises partenaires (comme Adobe) propose de lier **les informations d'identification** au contenu. Il est par exemple possible d'activer dans adobe (en beta) le **Content Credentials** qui permet d'activer la vérification du contenu.

Par la suite, les partenaires seront plus larges et pourront même toucher les fabricants d'appareils photos afin d'avoir cet historique à la source mais aussi les outils d'IA génératives d'image.

Les types de fichier pris en compte :

Photo: JPEG, PNG, TIFF, HEIC

Autres: PDF

Audio: WAV, MP3, AAC

Vidéo: MP4, MOV, AVI

Fonctionnement du C2PA

Le C2PA utilise des **métadonnées** intégrées qui accompagnent chaque fichier numérique. Ces données comprennent des informations sur la **création,** les **modifications** et la **distribution** du contenu, permettant ainsi aux utilisateurs de retracer
l'historique du fichier.

Les étapes en détails :

1. **Capture du Contenu et Génération de Métadonnées**

Lorsque du contenu est créé (image, vidéo, audio), le système commence par générer des **métadonnées cryptographiques** qui contiennent des informations essentielles, telles que :

•**L'auteur** ou le créateur du contenu (par exemple, le nom de l'appareil ou le logiciel utilisé),
•**La date et l'heure de création** (horodatage),
•**La géolocalisation** (si disponible et pertinente),
•**L'appareil utilisé** (modèle de caméra, smartphone, logiciel de génération de contenu comme Adobe Photoshop ou un outil d'IA),
•**Informations spécifiques sur le processus** (par exemple, l'utilisation d'une IA générative).

2. Signature Numérique

Ces métadonnées sont ensuite **signées numériquement** à l'aide d'une clé cryptographique, ce qui permet de garantir leur intégrité. La **signature numérique** agit comme une empreinte unique, attestant que le contenu et ses métadonnées n'ont pas été altérés depuis leur création.

•Chaque modification ultérieure du contenu, telle que l'édition ou la retouche, génère une **nouvelle signature** et une mise à jour des métadonnées, de sorte que l'historique des changements soit préservé.

Fonctionnement du C2PA

3. Intégration des Métadonnées dans le Fichier

Les métadonnées signées sont **intégrées directement dans le fichier (Manifest)** numérique (comme une image JPEG, une vidéo MP4, etc.) ou associées au fichier à travers une référence externe (comme un lien). Cela signifie que le fichier lui-même contient des informations qui décrivent son origine et ses éventuelles modifications, créant ainsi une **chaîne de provenance**.

•Cette chaîne de provenance suit le contenu tout au long de son cycle de vie, enregistrant chaque modification, copie ou manipulation.

Image de structure du manifest -> 2 slides*

4. Publication et Partage

Une fois le contenu capturé et les métadonnées intégrées, celui-ci peut être **partagé** ou **publié** sur différentes plateformes (réseaux sociaux, sites web, etc.). Lors du partage, les **métadonnées cryptographiques** et les signatures numériques sont transmises avec le fichier.

•Les plateformes ou applications compatibles avec le standard **C2PA** peuvent lire et afficher ces métadonnées pour permettre à l'utilisateur de vérifier la provenance et l'intégrité du contenu.

Fonctionnement du C2PA

5. Vérification et Authentification

Lorsque le contenu est consulté par un utilisateur ou une plateforme, un processus de **vérification** est lancé pour confirmer que les métadonnées n'ont pas été altérées. Cela comprend :

- **Vérification de la signature numérique :**

L'utilisateur peut vérifier l'authenticité du contenu en validant la signature numérique. Si la signature est valide, cela prouve que le contenu et ses métadonnées sont intacts depuis leur création.

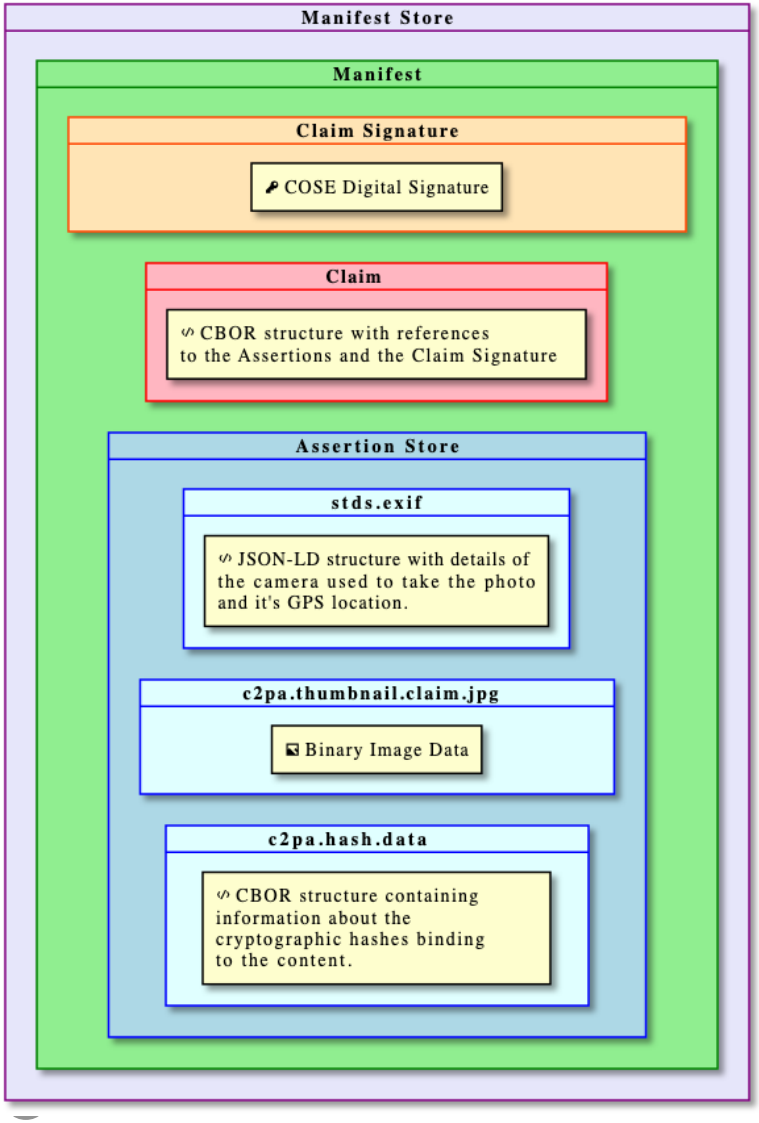
- **Consultation de l'historique de provenance :**

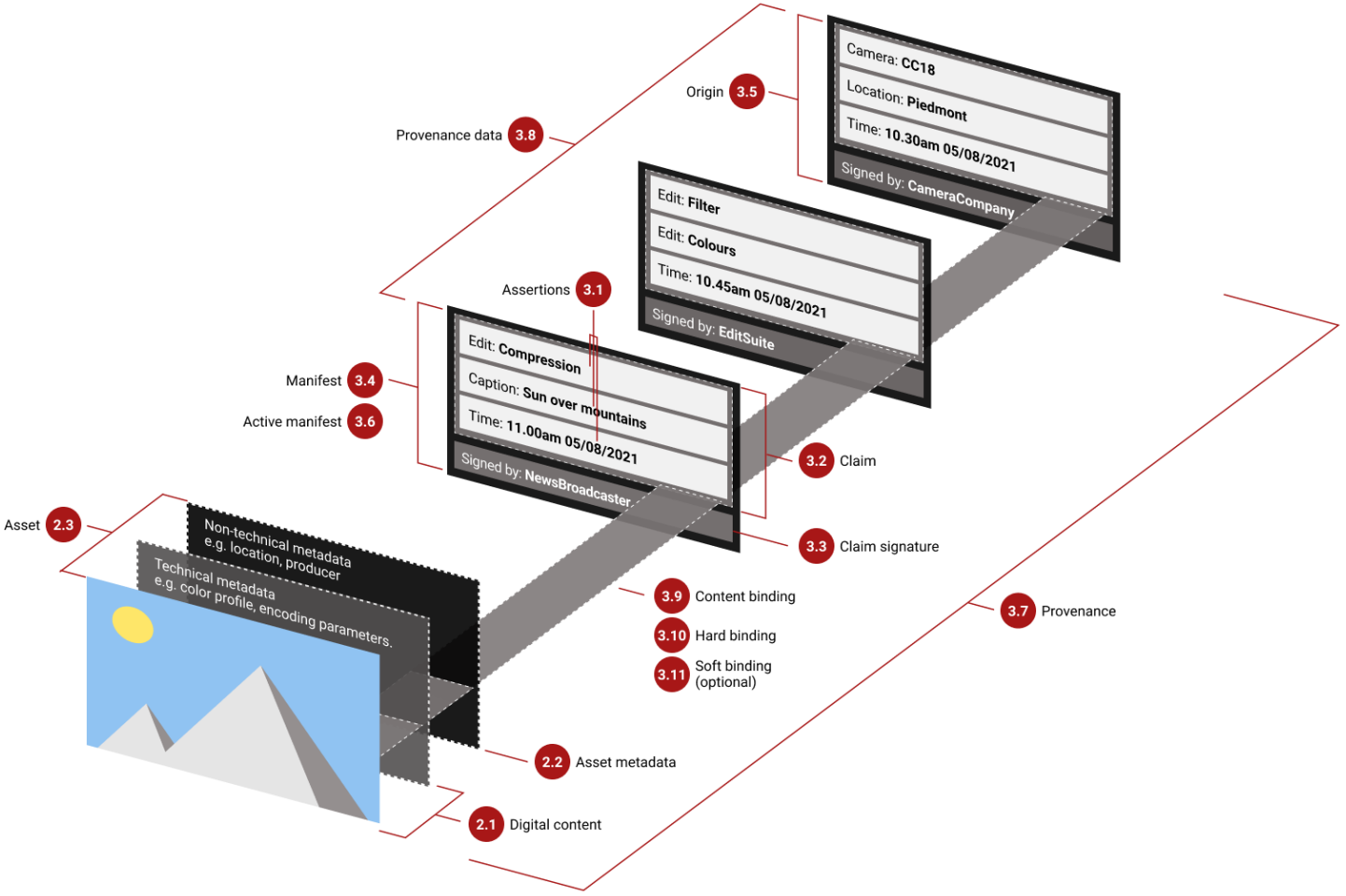
Les utilisateurs peuvent également consulter les **métadonnées d'origine** et les modifications apportées au contenu. Cela permet de voir **qui a créé le contenu**, quand et comment il a été modifié, ce qui est utile pour lutter contre les falsifications ou les manipulations.
Par exemple, si une image a été retouchée avec Photoshop, le fichier contiendra des informations détaillées sur la date de modification, l'auteur des retouches, et les outils utilisés.

- **Protection contre les altérations :**

Si quelqu'un tente de modifier le fichier sans mettre à jour les métadonnées ou sans suivre le processus de signature numérique, la **signature numérique devient invalide**. Cela permet aux utilisateurs et aux plateformes de détecter facilement si le contenu a été falsifié ou modifié de manière non autorisée.

Structure Manifest C2PA





Démonstration

Lien content credentials verify -> https://contentcredentials.org/verify

Lien image d’exemple C2PA -> https://c2pa.org/public-testfiles/image/

Avantages

- **Traçabilité** : Garantit la provenance et l'authenticité des contenus numériques (créateur, appareil, modifications).
- **Lutte contre la désinformation** : Aide à détecter et prévenir les deepfakes et contenus manipulés.
- **Soutien des grands acteurs** : Adobe, Microsoft, Intel, et d'autres soutiennent ce standard ouvert, favorisant son adoption.
- **Interopérabilité** : S’applique à divers types de fichiers (images, vidéos, audio) et secteurs.
- **Sécurité accrue** : Signature cryptographique garantissant l'intégrité des métadonnées.

Inconvénients

- **Adoption inégale** : Pas encore pris en charge par toutes les plateformes, ce qui limite l'usage.
- **Charge supplémentaire pour les créateurs** : Ajoute des étapes au flux de travail des créateurs.
- **Problèmes de confidentialité** : Risque de divulgation de données sensibles (géolocalisation, informations personnelles).
- **Impact sur la performance** : L'ajout de métadonnées peut augmenter la taille et la complexité des fichiers.

Conclusion & Questions ?

Le standard **C2PA** (Coalition for Content Provenance and Authenticity) offre une solution innovante pour garantir la provenance, l'authenticité et l'intégrité des contenus numériques grâce à des métadonnées sécurisées par des signatures cryptographiques. Il contribue à lutter contre les manipulations, comme les deepfakes, et renforce la confiance dans les fichiers partagés en ligne. Bien qu'il présente des défis techniques et nécessite une adoption plus large, C2PA est un outil essentiel pour assurer la transparence dans l'écosystème numérique moderne.

Merci de m’avoir écouté ☺