Towards an international standard to establish trust in media production, distribution and consumption

Frederik Temmermans
Department of Electronics & Informatics (ETRO)
Vrije Universiteit Brussel and imec
Brussels / Leuven, Belgium
frederik.temmermans@vub.be

Sabrina Caldwell
College of Engineering, Computing and Cybernetics
The Australian National University
City, Australia
sabrina.caldwell@anu.edu.au

Symeon Papadopoulos
Media Analysis, Verification and Retrieval Group
Centre for Research and Technology Hellas
Thermi, Thessaloniki, Greece
papadop@iti.gr

Fernando Pereira
Instituto Superior Técnico
Instituto de Telecomunicações
Lisboa, Portugal
fp@lx.it.pt

Philippe Rixhon
Centre for Blockchain Technologies
University College London
London, United Kingdom
philippe@rixhon.net

***Abstract*—Advances in media content manipulation and artificially generated content pose new challenges to the assessment of media authenticity. While automated detection methods can provide meaningful insights and decision support in some scenarios, they cannot provide trustworthy and comprehensive information about the origin and provenance of media assets. Therefore, a longer-term approach should rather focus on secure and interoperable annotations related to the creation and provenance of media. In October 2020, the JPEG Committee initiated a standardization exploration named "JPEG Fake Media" to address these needs. Subsequently, since many of the requirements, for example related to secure annotation and identification of media assets, are also relevant to achieve interoperability in Non-Fungible Tokens (NFTs) an additional exploration was initiated, specifically focused on standardization needs for NFTs. In April 2022 a first Call for Proposals on JPEG Fake Media was issued. Based on the responses to the call, a new standardization project named JPEG Trust was initiated to specify an interoperable framework for establishing trust in media production, distribution, and consumption. This paper presents the journey of JPEG to leverage formal methods of standardization in this context, starting from the initial JPEG Fake Media exploration, followed by the subsequent consideration of NFT use cases and requirements, through to the commencement of the new JPEG Trust international standard.**

***Index Terms*—media authenticity, interoperability, standardization, NFTs**

I. Introduction

Automatic assessment of media authenticity has traditionally been carried out with the help of media forensics methods, with different methods being proposed per modality, e.g. image [1], audio [2], video [3]. A wide range of methods have been proposed ranging from methods for identifying the camera model used to capture a photo to methods for localizing splicing and copy move forgeries. Yet, despite the breadth of media forensics methods, and their effectiveness in certain media authenticity assessment scenarios, it is widely acknowledged that it is very challenging to draw reliable conclusions in several *in the wild* settings, when little is known about the provenance of a specific media item [4]. Moreover, interpreting the results of media forensics methods is notoriously complicated and calls for subject matter experts.

The challenge is further exacerbated by the recent advent and rapid improvement of AI-based synthetic media, including deepfakes, Generative Adversarial Networks (GANs), Neural Radiance Fields and Diffusion Models. These models offer a variety of tools for either manipulating existing media (e.g., face swapping, editing expressions, changing the artistic style) or generating entire media assets from scratch (e.g. using random seeds or using text prompts). While there is a rapidly growing body of research dedicated to the detection of synthetic manipulations, commonly referred to as deepfake detection [5]–[7], there are similar challenges, as in the case of media forensics analysis, when trying to assess the authenticity of media and to establish whether generative AI was involved in the process of their generation or processing.

Even in cases where one is able to tell that a media asset has been manipulated or synthesized using AI models, there is often high uncertainty with respect to the particular method/model used to do so; it is almost infeasible to draw conclusions when a variety of tools and manipulation methods are used in combination. Last but not least, synthetic media detection methods often suffer from a high number of false positives, which gives rise to the issue of the *liar's dividend* [8], i.e. the plausible dismissal of a media asset as a deepfake (or manipulated), even in cases where the media asset in question is genuine.

To address the above challenges that are inherent to forensics- and AI-based detection approaches, it became clear

Contribution of Symeon Papadopoulos was done in context of the European Commission vera.ai HE-101070093 project. All authors of this papers are members of the JPEG Committee (ISO/IEC JTC 1/SC 29/WG 1) and were actively involved in the presented activities.

979-8-3503-3959-8/23/$31.00 ©2023 IEEE

that a new standard is needed to ensure the secure, interoperable and reliable annotation of media modifications. To this end, the Joint Photographic Experts Group (JPEG, ISO/IEC JTC 1/SC 29/WG 1) initiated an exploration to address standardization needs related to media authenticity under the name "JPEG Fake Media". Since the committee identified several overlaps with respect to requirements related to Non-Fungible Tokens (NFTs), an additional exploration was initiated to specifically focus on use cases and requirements in this domain, while keeping both activities clearly aligned. The latest step forward is the launch of a new standardization project named "JPEG Trust". Initially, JPEG Trust will address the core requirements related to media authenticity as identified in the JPEG Fake Media exploration. In the future, the project will be further extended to cover additional trust related topics, including the outcomes of the exploration on NFTs.

This paper describes the journey of JPEG to leverage formal methods of standardization to grapple with these complex problems, and undertake the development of a new standard to establish trust in media production, distribution, and consumption. The following sections of this paper are organized as follows. Section II describes the general principles followed by JPEG to establish new standards. Sections III and IV describe the JPEG Fake Media and NFT explorations respectively. Finally, Section V briefly introduces the upcoming JPEG Trust standard and Section VI provides closing remarks and a glimpse on the future of the standard.

II. JPEG Standardization

The main goal of JPEG standards is to facilitate the creation of interoperable media-based ecosystems, notably involving imaging modalities [9]. While coding is a familiar technology requiring a standard specification to support interoperability, the growing sophistication of emerging media-based applications and services also requires the standard specification of more technologies, notably metadata, and other system-level tools [10].

To specify the most effective standards, JPEG standardization starts with an exploration phase followed by a final specification phase which typically involves competitive and collaborative efforts.

The main goal of the exploration phase is to understand in detail the problems and needs in a specific domain requiring JPEG standardization attention, and to describe them in terms of use cases and requirements. While the use cases correspond to practical situations where the standardization needs are clear, the requirements translate those needs into technical requests to be addressed by the potential standard. This exploration process must happen in intense interaction with the relevant stakeholders to collect solid and meaningful use cases and requirements. This engagement is commonly conducted in multiple workshops focusing on the technological challenges, as well as legal implications, societal impact and end-user needs.

After the exploration phase, based on the collected use cases and requirements, and its proximity to JPEG competences, JPEG may decide to initiate a standardization project. At this stage, a more formal process starts, typically signaled with the launching of a competitive Call for Proposals asking the 'outside world' for technical contributions addressing the previously identified requirements. These contributions are carefully analyzed against the requirements to select the initial set of tools that will then be collaboratively improved, completed and optimized towards the final standard specification.

In the specific case of the JPEG Trust standard, which is discussed in more detail later in this paper, there were two preceding explorations. The first exploration, JPEG Fake Media, is finalized and the results of the Call for Proposals are being incorporated in the first part of JPEG Trust. At the time of writing, the second exploration, JPEG NFT, is still ongoing. It is expected that the outcomes of JPEG NFT will further extend the JPEG Trust standard in the future.

III. JPEG Fake Media Exploration

A. Scope and approach

The JPEG Fake Media exploration was initiated by the JPEG Committee in October 2020 with the aim to explore standardization needs related to fake media and media authenticity. The scope was defined as *"the creation of a standard that can facilitate a secure and reliable annotation of media asset creation and modifications. The standard shall address use cases that are in good faith as well as those with malicious intent."* [11]. To engage with stakeholders and identify relevant use cases, several workshops were organized. The presentations and recordings of these workshops are available on the JPEG website [12].

B. Use cases and requirements

Perhaps not surprisingly in the growing climate of concern over the social impacts of misinformation and disinformation, initial use cases focussed on 'fake media' and the lack of presence and security of claims of ownership and veracity of images. During this phase, concerns about inauthentic images have only been exacerbated by the disturbing verisimilitude of deepfake video production, and AI based image synthesis models such as Generative Adversarial Networks (GANs) and Stable Diffusion that are capable of mimicking reality. Furthermore, new questions about old images have been raised, such as the debate as to whether dementia research progress was retarded by manipulated medical images in an influential 2006 Nature article [13], or whether traditional problems with adolescent children's body images caused by manipulated images of models have simply gone online through social media platforms [14].

However, equally important in tackling fake media, or trust in images, is recognition that media modifications are often positive and even necessary and normal elements of producing images. These use cases are just as dependent on interoperable standards for annotating images with modification and provenance metadata as fake media prevention use cases. Wedding photographers routinely optimize their photographs to produce the glow of a special day for their customers. Creative artists use their special talents to produce images (photographic or otherwise).

The use cases considered by the working group ultimately coalesced around the areas of mis- and disinformation, forgery and media forensics, media creation and media modification. Media creators, end users and consumers needing the proposed secure media annotation of JPEG Trust are myriad, for example including professional photographers, amateur scientists, genealogists, insurance companies, news agencies, as well as casual consumers of social media.

Such a wide range of application domains necessitated a broad and flexible range of requirements, which represented the next step in building the platform for a robust international standard.

Based on the identified use cases, an extensive list of requirements was extracted which are documented in the Use Cases and Requirements for JPEG Fake Media document [11]. The requirements were grouped as follows:

- Media creation and modification descriptions
- Metadata embedding and referencing
- Authenticity, integrity, and trust model

Upon completion of this part of the exploration, it was determined that the technical requirements were capable of being addressed within the current affordances of technology, and that standardization was justified. Simultaneously, additional aspects of media asset identification and authenticity arose, most notably in the rise of Non-Fungible Tokens (NFTs).

IV. JPEG NFT Exploration

A. Scope and approach

Non-Fungible Tokens (NFTs) are digital certificates authenticating properties of an asset and using blockchain technology. Recently, NFTs have garnered considerable interest. Several digital assets that NFTs point to are either in existing JPEG formats or can be represented in current and emerging formats under development by the JPEG Committee. However, various provenance and security concerns have been raised about NFTs. To better understand requirements for media formats, the JPEG Committee launched the JPEG NFT exploration. The scope was defined as *"the creation of effective specifications that support a wide range of applications relying on NFTs applied to media assets. The standard shall be secure and eco-friendly, allowing for an interoperable ecosystem relying on trustworthy NFTs within a single application or across applications."*. The committee strives to engage stakeholders from diverse backgrounds, including technical, legal, artistic, and end-user, to establish use cases and requirements. At the time of writing, the latest "Use Cases and Requirements for JPEG NFT" document was released in January 2023 [15].

B. Use cases and requirements

Numerous applications have already benefited from the specific properties of NFTs applied to physical, digital, and phygital assets, while others have been predicted but not thoroughly explored. A non-exhaustive list of such use cases comprises intellectual property, trading of artworks and collectibles, gaming, ticketing, and academic certification.

Each NFT includes a unique identifier and metadata about the related asset. *Minting* an NFT refers to generating the cryptographic token used to represent a unique asset and recording it on a blockchain. Once *minted*, an NFT cannot be swapped or edited, making it ideal for establishing and tracking provenance, ownership or access rights of natively digital assets or tokenized physical goods. Smart contract programming facilitates and records transfers of NFTs between buyers and sellers.

Ownership of an NFT does not inherently grant copyright to whatever digital or physical asset the token represents. According to legal scholar Rebecca Tushnet, "From an IP perspective, NFTs do not change anything. If you did not have the rights to distribute a work before, you do not have them now. If the sale or memorialization of an NFT involves reproducing and distributing a work that is under copyright, then copyright will cover those reproductions unless a limitation or exception like *fair use* applies." [16].

NFTs themselves can be subject to IP protections, including copyright, design, patent, and trademark rights. As such, NFT purchasers should pay attention to what IP rights, if any, come part and parcel with the NFT.

Based on the identified use cases, an initial set of JPEG NFT requirements [15] have been identified and organized in the following main categories:

- Metadata descriptions
- Metadata embedding and referencing
- Provenance, authenticity, and integrity
- Media asset registration format

From these requirements categories it becomes apparent that many of the underlying needs overlap with those of JPEG Fake Media introduced in the previous section. Therefore, it is utterly important that the subsequent specifications are aligned with each other. Moreover, there is also an overlap in scope. Trustworthy knowledge about the authenticity of assets is crucial when transferring ownership. Contrariwise, in some cases blockchain technology can serve as one of the tools to establish trust related to the authenticity of the asset.

V. JPEG Trust Standard

A. Scope and approach

The standardization project JPEG Trust was initiated after completion of the Final Call for Proposals on JPEG Fake Media [17]. The choice for the name *JPEG Trust* was signaled early in the work in acknowledgement of the importance of the solution for not just addressing fake media, but also valid image modifications as well as new aspects of image consumption, including NFTs. The standardization project was renamed JPEG Trust to reflect the expected benefits of the standard, which are to instill trust, interoperability and increased utility of the media assets we create and use every day.

The scope of JPEG Trust is defined as *"The scope of JPEG Trust is to provide a framework for establishing trust in media.*

This framework includes aspects of authenticity, provenance and integrity through secure and reliable annotation of the media assets throughout their life cycle". While initially the project focuses on covering the JPEG Fake Media use cases and requirements, it is expected that the project will be further extended in the future, including tools to achieve improved interoperability in the domain of NFTs.

As a first step, the first Part of JPEG Trust named *Core Foundation*, focuses on the core building blocks to build a framework to establish trust in media. This part builds on the outcomes of the exploration on JPEG Fake Media, introduced in Section III. More specifically, it builds on the responses to the "Final Call for Proposals on JPEG Fake Media" that was issued in April 2022 [17]. In response to this call, in October 2022, the committee received the following 6 proposals:

1) Adobe/C2PA: C2PA Specification
2) Huawei: Provenance and Right Management for Digital Contents in JPEG Fake Media
3) Sony Group Corporation: Methods to keep track provenance of media asset and signing data
4) VUB/imec: Improved media revision history tracking via asset decomposition and serialization
5) UPC: MIPAMS Provenance module
6) Newcastle University: TRusted mediA dIstribuTion (TRAIT)

An extensive evaluation of these proposals was conducted according to the following basic principles:

- the process was open, transparent, fair and unbiased
- the process allowed for detailed and deep technical discussion
- the process was efficient and timely towards the goal, i.e. defining the starting solution for the Fake Media standard

As a result of the evaluation phase it was decided to initiate the new JPEG Trust standardization leveraging upon several elements of the various proposals as well as the JPEG ecosystem of widely adopted standards. While still under subject to changes, the Core Foundation will address the following topics:

- JPEG Trust framework
- Media asset life cycle annotations
- Embedding and referencing
- Privacy and protection
- Identification of assets and actors
- Media asset authenticity & integrity
- Trust indicators

In addition, several informative annexes will depict the usage of the framework in several usage scenarios and domains, such as representative photography and GLAM (Galleries, Libraries, Archives and Museums).

B. Next steps and timeline

Currently a first working draft of Part 1 of JPEG Trust, Core Foundation, is under development. It is expected that the work will be published as an international standard in 2025. Over time new Parts will be added to cover additional trust related topics. These topics may include NFTs as well as media asset registration and version control. Specifically for JPEG NFT the release of a Final Call for Proposals is currently scheduled for July 2023.

VI. Closing remarks

It is expected that JPEG Trust will become an international standard in 2025. The standard will provide a framework for establishing trust in media production, distribution and consumption. The framework provides important aspects of authenticity, provenance and integrity through secure and reliable annotation of the media assets throughout their life cycle. As a first step, the current foundational phase of the standard focuses on structuring the core building blocks to build a framework to establish trust in media. More specifically, this first part mainly focuses on covering the use cases and requirements identified during the JPEG Fake Media exploration. It is however expected that the standard will be further extended in the future. One of the expected extensions will focus specifically on tools to establish interoperability for NFTs. The required building blocks will be shared as much as possible with the core foundation to achieve optimal interoperability. In addition to the focus on NFTs, more elaborated specifications for media asset registration and version control are on the radar.

References

[1] A. Piva, "An overview on image forensics," *ISRN Signal Processing*, vol. 2013, p. 496701, Jan 2013. [Online]. Available: https://doi.org/10.1155/2013/496701

[2] M. Zakariah, M. K. Khan, and H. Malik, "Digital multimedia audio forensics: past, present and future," *Multimedia Tools and Applications*, vol. 77, no. 1, pp. 1009–1040, Jan 2018. [Online]. Available: https://doi.org/10.1007/s11042-016-4277-2

[3] S. Milani, M. Fontani, P. Bestagini, M. Barni, A. Piva, M. Tagliasacchi, and S. Tubaro, "An overview on video forensics," *APSIPA Transactions on Signal and Information Processing*, vol. 1, p. e2, 2012.

[4] M. Zampoglou, S. Papadopoulos, and Y. Kompatsiaris, "Large-scale evaluation of splicing localization algorithms for web images," *Multimedia Tools and Applications*, vol. 76, no. 4, pp. 4801–4834, Feb 2017. [Online]. Available: https://doi.org/10.1007/s11042-016-3795-2

[5] Y. Mirsky and W. Lee, "The creation and detection of deepfakes: A survey," *CoRR*, vol. abs/2004.11138, 2020. [Online]. Available: https://arxiv.org/abs/2004.11138

[6] R. Tolosana, R. Vera-Rodríguez, J. Fiérrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: A survey of face manipulation and fake detection," *CoRR*, vol. abs/2001.00179, 2020. [Online]. Available: http://arxiv.org/abs/2001.00179

[7] L. Verdoliva, "Media forensics and deepfakes: an overview," *CoRR*, vol. abs/2001.06564, 2020. [Online]. Available: https://arxiv.org/abs/2001.06564

[8] R. Chesney and D. K. Citron, "Deep fakes: A looming challenge for privacy, democracy, and national security," *SSRN Electron. J.*, 2018.

[9] S. Foessel, J. Ascenso, L. A. da Silva Cruz, T. Ebrahimi, P.-A. Lemieux, C. Pagliari, A. M. G. Pinheiro, J. Sneyers, and F. Temmermanns, "Jpeg status and progress report 2022," *SMPTE Motion Imaging Journal*, vol. 131, no. 8, pp. 111–119, 2022.

[10] F. Temmermans, A. Kuzma, S. Choi, and P. Schelkens, "Adopting the JPEG systems layer to create interoperable imaging ecosystems," in *Optics, Photonics and Digital Technologies for Imaging Applications VI*, P. Schelkens and T. Kozacki, Eds., vol. 11353, International Society for Optics and Photonics. SPIE, 2020, p. 113530T. [Online]. Available: https://doi.org/10.1117/12.2557809

[11] JPEG, "Use Cases and Requirements for JPEG Fake Media," ISO/IEC JTC1/SC29 WG1, Tech. Rep. WG1N100156, 2022.

[12] "JPEG.org - Documentation on JPEG Fake Media," https://jpeg.org/jpegfakemedia/documentation.html, accessed: 2023-03-17.

[13] C. Pelc, "Alzheimer's study controversy: Where do we go from here? — medicalnewstoday.com," https://www.medicalnewstoday.com/articles/alzheimers-study-controversy-what-does-it-mean-for-future-research, [Accessed 17-Mar-2023].

[14] M. Kleemans, S. Daalmans, I. Carbaat, and D. Anschütz, "Picture perfect: The direct effect of manipulated instagram photos on body image in adolescent girls," *Media Psychology*, vol. 21, no. 1, pp. 93–110, 2018. [Online]. Available: https://doi.org/10.1080/15213269.2016.1257392

[15] JPEG, "Use Cases and Requirements for JPEG NFT v3.0," ISO/IEC JTC1/SC29 WG1, Tech. Rep. WG1N100391, 2023.

[16] C. Majocha, "MMemes for Sale? Making sense of NFTs. Harvard Law Today," https://hls.harvard.edu/today/memes-for-sale-making-sense-of-nfts/, 2021, [Accessed 19-Mar-2021].

[17] JPEG, "Final Call for Proposal for JPEG Fake Media," ISO/IEC JTC1/SC29 WG1, Tech. Rep. WG1N100157, 2022.