

**XIX** encontro nacional  
de pesquisa em  
ENANCIB ciência da informação

// SUJEITO INFORMACIONAL E AS  
PERSPECTIVAS ATUAIS EM CIÊNCIA  
DA INFORMAÇÃO. //

**22-26**  
**OUTUBRO**  
**2018**  
LONDRINA/PR



**XIX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XIX ENANCIB**

**GT ESPECIAL**

**POSSIBILIDADE DE IDENTIFICAÇÃO DE VIOLAÇÕES A DIREITOS AUTORAIS COM BASE EM METADADOS GERADOS NA BLOCKCHAIN: AVALIAÇÃO DA PLATAFORMA ORIGINAL.MY**

**José Eduardo Santarém Segundo**

**José Antonio Maurílio Milagre de Oliveira**

**THE POSSIBILITY OF IDENTIFYING COPYRIGHT VIOLATIONS BASED ON BLOCKCHAIN GENERATED METADATES: EVALUATION OF THE ORIGINAL.MY PLATFORM**

**Modalidade de apresentação: Pôster**

**Resumo:** A infraestrutura Blockchain, projetada em 2008, vem sendo considerada uma estrutura segura e descentralizada para registro de transações envolvendo ativos digitais ou criptomoedas. Sendo considerada um grande “livro razão”, permitindo que fatos, atos, registros e informações sejam eternizadas de forma indelével e autenticada por pontos conectados à rede, vem hoje sendo analisada como plataforma para outras funções, incluindo para registros de conteúdos protegidos por direitos autorais. O propósito do artigo é avaliar, a partir da Blockchain, quais metadados são gerados e que podem ser úteis na detecção de violação ou identificação de direitos autorais, com o registro de um Livro na aplicação Original.My. Como resultados, identificou-se que somente os metadados gerados não serão suficientes para detecção de cópias parciais e derivações, sendo proposto, para trabalhos futuros, a especificação de um formato de metadados que permita maior eficiência na detecção de cópias e violações.

**Palavras-Chave.** metadados, blockchain, direitos autorais, criptografia, direitos.

**Abstract:** The Blockchain infrastructure, designed in 2008, has been considered a secure and decentralized structure for recording transactions involving digital or cryptocurrencies assets. Being considered a great "ledger", allowing facts, acts, records and information to be eternally indelible and authenticated by points connected to the network, it is now being analyzed as a platform for other functions, including for registration of copyrighted content. The purpose of the article is to evaluate, from Blockchain, what metadata are generated and which may be useful in detecting copyright infringement or identification, with registration of a Book in the Original.My application. As results, it

was identified that only the generated metadata will not be enough to detect partial copies and derivations, and it is proposed, for future work, the specification of a metadata format that allows greater efficiency in the detection of copies and violations.

**Keywords.** metadata, blockchain, copyright, encryption, rights.

## 1 INTRODUÇÃO

O artigo tem por finalidade analisar quais metadados são gerados e gravados na *Blockchain* e que podem servir como recursos para aplicações na identificação e rastreamento de violações à direitos autorais.

A garantia de direitos autorais tem-se tornado um desafio com o desenvolvimento da Internet, tendo em vista a facilidade de compartilhamento de dados, com o livre trânsito de textos, fotos e vídeos, muitos deles, de forma que seus criadores não sejam remunerados pelas cópias das obras (POYATOS, 2018).

Entende-se a *Blockchain* como uma grande corrente, onde cada elo caracteriza uma transação ou um registro. Esta composição de elos é automatizada e não pode ser modificada sem “detecção”, sendo que as transações não podem ser apagadas (LEÃO, CANEDO, GOMES, 2018, p.1).

O presente trabalho está organizado da seguinte forma. Na seção 2 são contextualizados os metadados e como a estrutura *Blockchain* pode contribuir para rastrear informações e documentos na rede. Na seção 3 apresenta-se os resultados do registro de uma obra na *Blockchain*, partir da aplicação Original.My e quais metadados foram extraídos. Na seção 4, apresenta-se quais os resultados identificados sobre o registro e consulta a obra na *Blockchain*, limitações identificadas e uma proposta para melhoria dos processos de detecção de violações autorais. Na Seção 5 apresenta-se as conclusões do trabalho e propostas para trabalhos futuros.

## 2 OS METADADOS, A BLOCKCHAIN E SUA IMPORTÂNCIA NA IDENTIFICAÇÃO DE CONTEÚDOS PROTEGIDOS POR DIREITO AUTORAL

A *Blockchain* é considerada hoje um livro contábil (*ledger*), que registra transações de informações e ativos digitais. É possível armazenar em ordem cronológica transações, as tornando imutáveis. Este “livro” pode ser auditado a qualquer momento por sistemas, inclusive de registro de obras e conteúdos protegidos por direitos autorais.

Um artigo publicado por Nakamoto (2008) descreve pela primeira vez o funcionamento da *Blockchain*, tendo como suas características, a) ser *peer-to-peer*, onde as relações se dão sem intermediários, b) inexistir autoridade central que valide as transações, c) *proof-of-work*, ou a característica do esforço computacional em dificultar adulterações, d) sincronização, sendo que qualquer participante poderá se desligar da rede e voltar a qualquer momento, tendo que aceitar o bloco encadeado de transações, e) consenso entre a maioria, onde as transações serão inseridas na corrente se houver aprovação pela maioria dos pontos presentes na rede, através dos algoritmos de consenso. (LEÃO, CANEDO, GOMES, 2018, p.4).

Assim, qualquer conteúdo digital pode ter uma assinatura registrada na plataforma. Inicialmente se extrai a função matemática de uma frase, algarismo, documento, livro, foto ou obra, o que se denomina de *hash*, e a partir deste ponto registra-se esse *hash* na rede *Blockchain*, o qual passará a existir para sempre nesta rede, com a data e horário que o registro foi realizado, permitindo-se a consulta a qualquer momento. Considerando que o *hash* é um identificador único, somente um arquivo poderia ser gerado com aquele número e registrado na plataforma.

Após, é possível verificar o registro através do site <https://www.blockchain.com/pt/explorer>, por meio da digitação do código da transação.

Deste modo, as transações pendentes são agrupadas em um bloco por um dos nós da rede, considerando-se o bloco como a estrutura que custodia as transações a serem incluídas na *Blockchain*. Esse bloco possui cabeçalho que por sua vez registra metadados sobre o bloco, lista de transações válidas e o código *hash* do bloco anterior e seu próprio *hash* (Morabito, 2017). É neste cenário que ocorre o envio do bloco à rede sendo que os nós validam a transação, ocorrendo o fechamento do elo da corrente e sendo registrado de forma inviolável, com o acréscimo de um novo bloco ao final da corrente existente (Nakamoto, 2008), com o registro cronológico da transação.

Metadados são considerados “dados sobre dados” com o escopo de descrever a informação para posterior recuperação. Enquanto uma tentativa de padronização do tratamento dos recursos informacionais na web, os metadados constituem um foco de estudos bastante contextualizado dentro da Ciência da Informação (LOURENÇO, 2007, p.66).

E é neste contexto que os metadados são reconhecidos por Lourenço (2007, p. 67) como chave para resolução de problemas na recuperação da web, com a concepção de padrões únicos ou padrões específicos para cada necessidade e que sejam comunicáveis.

Como se vê, descrever e representar um recurso informacional de forma unívoca se torna uma constante (YAMANE e CASTRO, 2018, p.1) e deste modo conceitua-se metadados como “um conjunto de elementos que descrevem as informações contidas em um recurso, com o objetivo de possibilitar sua busca e recuperação” (GRÁCIO, 2002, p.23). Deste modo, resta evidente que metadados atrelados à obras e conteúdos protegidos por direitos autorais permitiriam não só descrever tais conteúdos, mas identificá-los, evitando-se ou detectando-se cópias.

Diante desta premissa, analisou-se no presente trabalho quais são os metadados gerados a partir do registro de um conteúdo protegido por direito autoral na Blockchain, ao final os descrevendo e avaliando se constituem suficientes para recuperação e busca de obras copiadas, bem como as limitações existentes. Ao final, apresenta-se sugestão de melhorias para um futuro “formato de metadados” que possibilite o uso por recursos e aplicações de busca de contrafação, entendido por “formato de metadados”, nos ensinamentos de Rossetto (2003,p.59), *"padrões que estabelecem regras para a definição de atributos (metadados) de recursos de informacionais, para a) obter coerência interna entre os elementos por meio de semântica e sintaxe; b) promover necessária facilidade para esses recursos serem recuperados pelos usuários; c) permitir a interoperabilidade dos recursos de informação"*.

### **3 ANÁLISE DE METADADOS GERADOS NA BLOCKCHAIN A PARTIR DO REGISTRO DE UM LIVRO**

A plataforma Original.my ([www.original.my](http://www.original.my)) é um sistema que interage com o código da *Blockchain*, coletando arquivos e artefatos digitais, e submetendo à registo no ambiente, permitindo verificações futuras. Para fins desta pesquisa, utilizou-se o serviço “*Bit Registry*”, que prova a autenticidade de documentos digitais.

Podem ser protegidos conteúdos como obras de arte, declarações, propostas, relatórios e qualquer outro tipo de documento. No entanto, a plataforma não nomina o serviço como “proteção autoral”, mas como “prova de autenticidade”, considerando que o documento recebe um “carimbo do tempo”.

Acessando-se a plataforma, inicialmente realizou-se o upload de um livro, arquivo em formato pdf, com 592 Kilobytes (KBs), o que consumiu em torno de 40 (quarenta minutos) para ser processado, com a informação “*Aguardando confirmação na Blockchain*” (MILAGRE, 2018).

Após esta fase, fora possível identificar os seguintes metadados gerados, relativos ao registro da obra na *Blockchain*:

**a) Assinatura digital do documento:**

505b994c26311e9dad33b2d0db888426633bd159920ba3c781ba633bf07c3501;

**b) Data e hora da assinatura digital:** 02-03-2018 11:33:37 BRT;

**c) Data do envio da transação à Blockchain:** DCR 02-03-2018 11:00:00 BRT;

**d) Data do registro na Blockchain:** DCR 02-03-2018 12:00:22 BRT;

**e) Código da transação:**

82f57c991dab62b81487cf8a56fc5deb4eb1d58c7530ebc273bd4fd0bcbcab8e;

**f) Certificado digital da transação:** arquivo remetido por e-mail.

A partir destes registros, recebemos um link onde através do da assinatura digital do documento (*hash*) foi possível verificar o certificado do registro junto à Blockchain: <https://originalmy.com/bitregistro/505b994c26311e9dad33b2d0db888426633bd159920ba3c781ba633bf07c3501>

Na plataforma Original.my é possível ainda verificar se um arquivo está ou não na *Blockchain*, sendo que o arquivo é a própria chave. É realizado o cálculo do *hash* do arquivo e a pesquisa por *hash* idêntico na plataforma blockchain. Em nossos testes, em 24 de julho de 2018, submetemos o livro com algumas derivações no conteúdo e o resultado foi que o arquivo (livro eletrônico) não estava certificado no Blockchain.

Apurou-se que uma pequena alteração no arquivo de entrada, produzirá a completa modificação do arquivo de saída, ou na geração do *hash* (SCOTT, 2016, p.1.). Deste modo, uma pequena alteração no conteúdo da obra é suficiente alterar o valor do metadado *hash*, driblando a plataforma *Blockchain*, permitindo o registro de obra parcialmente modificada ou com pequenas derivações. Como se sabe, a função do *hash* criptográfico é coletar entradas e calcular uma saída em forma de código, sendo que para entradas diferentes, saídas serão diferentes.

#### **4 RESULTADOS IDENTIFICADOS E PROPOSTAS PARA MELHORIA**

Não se discute que metadados sempre existiram ou puderam ser gerados em ou a partir de documentos digitais, porém, com a *Blockchain* segura e imutável, temos uma mecanismo descentralizado que também armazena os metadados que acompanham um

arquivo, logo, sendo possível o desenvolvimento de aplicações que façam a checagem, a exemplo, plataformas de submissões de textos e artigos científicos, capazes de identificar obras já submetidas.

Assim, os metadados gerados podem ser usados por recursos de rastreabilidade ou aplicações que proporcionarão a checagem de registros de titularidade de obras, informações e bancos de dados, o que poderá ser utilizado em disputas.

Embora seja possível rastrear o registro de trabalhos por meio de metadados que são gerados e que se integram à Blockchain, identificou-se que derivação de textos, variação ou cópia de trechos pode não ser detectada, já que o código matemático (*hash*) é gerado a partir do documento todo e neste caso, o *hash* de um conteúdo contrafeito e modificado seria distinto. Do mesmo modo, os recursos de rastreamento não poderão considerar apenas a checagem da assinatura do arquivo (*hash*), considerando que trechos podem ter sido copiados, ou mesmo o arquivo ter sido inserido em outro, o que obstaculizaria a detecção. Como explica Cameron-Huff, *apud* Jessie Willms (2016) “*Um hash de um trabalho não é o mesmo que o trabalho em si e os direitos autorais geralmente são mais amplos do que apenas o arranjo específico dos bits*”

Uma forma de tangenciar esta dificuldade é lecionada por SCOTT (2016, p.3) onde ao invés de da geração de um *hash* de uma informação ou obra completa, se buscaria o hash de um número na informação (de 4 dígitos), o que geraria um hash menos extenso, gerando maior possibilidade de identificação de um conteúdo derivado.

Outra proposta é apresentada pela aplicação *StopTheFakes.io*. A aplicação utiliza a figura de um “denunciante”, denominado “*doer*”. O titular ou “*requestor*” *submete* a obra à *Blockchain* e solicita o interesse no rastreio autoral, sendo que a partir daí o “*doer*” pode procurar violações na internet e submeter ao “*requestor*”, *pela plataforma*, sendo remunerado por isso. A solução no entanto não automatiza a detecção da contrafação. Uma outra proposta para aprimorar o processo de rastreabilidade de direitos autorais que apontamos é a utilização da técnica “*fuzzy hashing*” nos arquivos submetidos à *Blockchain*, para geração de um metadado contendo alta ou baixa similaridade entre documentos comparados. Trata-se de uma técnica que permite identificar arquivos e conteúdos distintos, que possuem *hashes* totalmente distintos, mas quase idênticos ou muito similares. Com a proposta, seria possível, antes de um registro na *Blockchain*, submeter o arquivo análise fuzzy,

que geraria um “metadado de similaridade”, bloqueando-se ou sinalizando o registro de obras quase idênticas, com copia de trechos ou derivações, embora com hashes distintos.

A técnica de *fuzzy hashing* consegue “tratar” pequenas alterações nos arquivos, considerando que o algoritmo consegue extrair o *hash* de blocos do livro, texto, obra, etc., indicando similaridades que embora não fossem detectadas na *Blockchain*, que considera o código de todo o documento, são evidentes (RABELO, 2011).

Neste sentido, importa destacar em conclusão que “recursos” automatizados de detecção de violações autorais deverão considerar outros domínios, como análise de *big data*, para se verificar quantas pessoas estão utilizando um conteúdo digitalmente, inteligência artificial, para que se possa considerar não só a estrutura do arquivo, mas a partir do conteúdo, possa o mesmo ser “*taggeado*” com metadados adicionais, bem como “*web crawling*”, que poderá percorrer sites e varrer imagens e conteúdos, checando por licenças válidas. Somente os metadados identificados nesta pesquisa, gerados atualmente, a partir do registro de um livro em formato digital na *Blockchain*, não são eficientes para identificação automatizada e rastreamento de conteúdos protegidos por direitos autorais, quando ocorre uso parcial ou derivações, ainda que mínimas.

## 5 CONCLUSÕES E TRABALHOS FUTUROS

Com o presente trabalho, identificou-se que o mecanismo de registro de obras *Original.my*, que grava na *Blockchain*, gera 6 (seis) registros de metadados, que podem ser úteis à comprovação a titularidade de conteúdos, textos, fotos e vídeos. Por outro lado, restou evidenciado que o registro por si só não impede registros fraudados de obras derivadas, copiadas ou minimamente alteradas, considerando que o metadado (*hash*) pode ser modificado a partir de pequenas alterações na obra original, o que geraria a informação de que um arquivo não é cópia de um original já registrado. Apresentada uma proposta de aprimoramento da detecção de contrafação, de uso em recursos ou aplicações que busquem automatizar detecção de plágio ou violação a direitos autorais com base na *Blockchain*, com destaque para a técnica de “*fuzzy hashing*”, que pode gerar um metadado de similaridade e por conseguinte, identificar arquivos com alta semelhança, mesmo com *hashes* distintos.

Como trabalhos futuros, pretende-se avaliar outras plataformas, bem como desenvolver uma solução capaz de, considerando outras características que não somente o *hash* de um arquivo ou obra, possa pesquisar automaticamente por contrafações, cópias de

trechos e outras derivações ilegais ou não autorizadas envolvendo conteúdos protegidos por direitos autorais. Pretende-se igualmente apresentar uma proposta de formato de metadados com a finalidade específica de identificação e recuperação de informações relativas a conteúdos protegidos, que possam ser registrados na *Blockchain*.

## REFERÊNCIAS

BARRETO, Aldo de Albuquerque. Mudança estrutural no fluxo do conhecimento: a comunicação eletrônica. *Ciência da Informação*, Brasília, v.27, n.2, p.122-127, maio/ago. 1998.

GRÁCIO, J. C. A. Metadados para a descrição de recursos da Internet: o padrão Dublin Core, aplicações e a questão da interoperabilidade. 2002. 127 f. Dissertação (Mestrado em Ciência da Informação)–Faculdade de Filosofia e Ciências, Universidade Estadual Paulista,

Marília, 2002.

KODAKONE. **KodakOne Website**. 2018. Disponível em: <<https://kodakone.com/>>. Acesso em: 9 jun. 2018.

KOMBLUM, Jesse. **SSDEEP. Fuzzing Hashing Programa**. Disponível em:<<https://ssdeep-project.github.io/ssdeep/index.html>> Acesso em: 25 jul. 2018

LEÃO, Heloise Acco Tives; CANEDO, Edna Dias; GOMES, Alexandre Rodrigues. **Proposta de Uso do Blockchain para Validação de Documentos de Instituições de Ensino Superior**.

Disponível em: <<http://wpos.cic.unb.br/wpos2017/artigos/sess%C3%A3o%205/artigo-blockchain-resumido.pdf>> Acesso em: 01 ago. 2018.

LOURENÇO, Cíntia Azevedo. Metadados: O Grande desafio na organização da web. Disponível em:<<http://www.brapci.inf.br/index.php/article/download/12781>> Acesso em: 13 jul. 2018.

MILAGRE, José Antonio. Como fiz meu primeiro registro de livro na blockchains e protegi meus direitos autorais. Disponível em: <<http://josemilagre.com.br/blog/2018/03/16/como-fiz-meu-primeiro-registro-de-livro-na-blockchain-e-protegi-meus-direitos-autorais/>> Acesso em: 01 ago. 2018.

MORABITO, Vincenzo. **Blockchain value system**. Disponível em: <[https://link.springer.com/chapter/10.1007/978-3-319-48478-5\\_2](https://link.springer.com/chapter/10.1007/978-3-319-48478-5_2)> Acesso em: 09 jul. 2018.

NAKAMOTO, Satoshi. Bitcoin: **A Peer-to-Peer Eletronic Cash System**. Disponível em: <<https://bitcoin.org/bitcoin.pdf>> Acesso em: 05 jul. 2018.

ORIGINALMY.COM. Disponível em: <<https://originalmy.com/>>. Acesso em: 07 ago. 2018.



BRASIL. **Lei 9.610 de 1998. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências.** Disponível em:

<[http://www.planalto.gov.br/ccivil\\_03/Leis/L9610.htm](http://www.planalto.gov.br/ccivil_03/Leis/L9610.htm)>. Acesso em: 07 ago. 2018.

POYATOS, Henrique. **Blockchain e o novo “Kodak moment”.** Disponível

em:<<https://pt.linkedin.com/pulse/blockchain-e-o-novo-kodak-moment-henrique-poyatos>>. Acesso em: 07 ago. 2018.

RABELO, Luiz. **Fuzzy Hash sem mistérios.** Disponível em:

<<http://forensics.luizrabelo.com.br/2011/04/fuzzy-hash-sem-misterios.html>> Acesso em: 25 jul. 2018

ROSETTO, M. Metadados e formatos de metadados em sistemas de informação: caracterização e definição. 2003, 95 f. Dissertação (Mestrado em Ciências da Comunicação)– Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2003.

SCOTT, Michael. The Essence of the Blockchain. Disponível em:

<<https://www.miracl.com/press/the-essence-of-the-blockchain>> Acesso em: 07 ago. 2018.

STOP THE FAKES. **Blockchain service Anti-Counterfeit & Copyright Infringement.** Disponível em:<[https://stopthefakes.io/wp-content/uploads/2018/05/wp\\_en-1.pdf](https://stopthefakes.io/wp-content/uploads/2018/05/wp_en-1.pdf)> Acesso em: 04 ago. 2018

WILLMS, Jessie. **Is Blockchain-Powered Copyright Protection Possible?** Disponível

em:<<https://bitcoinmagazine.com/articles/is-blockchain-powered-copyright-protection-possible-1470758430/>> Acesso em: 07 ago. 2018.

YAMANE, Gabriela Aparecida da Cunha. O estudo e a identificação dos padrões de metadados para a representação e a recuperação da imagem digital na perspectiva da web. Disponível em:<<http://dx.doi.org/10.19132/1808-5245241.145-173>>. Acesso em: 10 jul.2018.