

A CADEIA DE BITS À LUZ DA AUTENTICIDADE: O CONTEÚDO BINÁRIO COMO ELEMENTO CENTRAL DA COMPOSIÇÃO DO DOCUMENTO ARQUIVÍSTICO NATO-DIGITAL

Natália Bruno Rabelo, Universidade Federal Fluminense (UFF), <https://orcid.org/0000-0002-8910-5516>

RESUMO

Este artigo dedica-se à análise dos fundamentos teóricos da Ciência da Computação acerca da arquitetura de computadores no escopo de transmissão do documento de arquivo nato-digital e da Arquivologia relacionados à autenticidade de documentos arquivísticos digitais. Este vínculo entre as duas áreas supracitadas origina a Arquivologia Computacional, visto que desde o final do século XX nota-se o crescimento da adoção e gestão de documentos arquivísticos digitais. Contudo, o tema arquitetura do documento de arquivo nato-digital ainda permanece objeto de estudo pouco explorado, onde demonstra-se a relação fundamental com o conteúdo binário que é apontado na literatura como vulnerável a perdas de dados de forma e conteúdo de acordo com eventos e, portanto, um desafio para a preservação digital no âmbito da Arquivologia. Logo, cabe refletir acerca da seguinte indagação: “qual a relação do conteúdo binário com a autenticidade do documento de arquivo nato-digital? Esta pesquisa de natureza qualitativa foi operacionalizada através de pesquisa bibliográfica de obras brasileiras e estrangeiras da Arquivologia e da Ciência da Computação que abordam, sequencialmente, documentos de arquivo digitais, autenticidade, arquitetura de computadores e aritmética binária. Os resultados sugeriram que não necessariamente a forma do documento está atrelada à autenticidade e que em razão do conteúdo binário estar sempre sujeito a alterações que ponham em risco sua integridade e simultaneamente ser o elemento representante da originalidade do documento, deve ser preservado em um ambiente seguro de adulterações para autenticações de cópia de um documento que tivera seu complemento digital comprometido.

Palavras-Chave: Arquivologia; Arquivologia Computacional; Documento Arquivístico Digital.

LA CADENA DE BITS A LA LUZ DE LA AUTENTICIDAD: EL CONTENIDO BINARIO COMO ELEMENTO CENTRAL DE LA COMPOSICIÓN DEL REGISTRO NATIVO-DIGITAL

RESUMEN

Este artículo está dedicado al análisis de los fundamentos teóricos de las Ciencias de la Computación sobre la arquitectura de computadoras en el ámbito de la transmisión del documento de archivo nativo-digital y la Ciencia de los Archivos relacionada con la autenticidad de los documentos de archivo digitales. Este vínculo entre las dos áreas antes mencionadas da origen a la Archivística Computacional, ya que desde finales del siglo XX ha habido un crecimiento en la adopción y gestión de documentos arquivísticos digitales. Sin embargo, la arquitectura del documento de archivo nativo-digital sigue siendo un objeto de estudio poco explorado, lo que demuestra la relación fundamental con el contenido binario que se señala en la literatura como vulnerable a la pérdida de datos de forma y contenido según los eventos, por lo tanto, un desafío para la preservación digital en el ámbito de la Archivística. Por lo tanto, vale la pena reflexionar sobre la siguiente pregunta: “¿cuál es la relación del contenido binario con la autenticidad del documento de archivo nativo-digital? Esta investigación cualitativa fue operacionalizada a través de la investigación bibliográfica de obras brasileñas y extranjeras de Archivística y Ciencias de la Computación que abordan secuencialmente los documentos de archivo digital, la autenticidad, la arquitectura informática y la aritmética binaria. Los

resultados sugirieron que la forma del documento no está necesariamente ligada a la autenticidad y que, debido a que el contenido binario siempre está sujeto a cambios que comprometen su integridad y, al mismo tiempo, es el elemento representativo de la originalidad del documento, debe ser preservado en un ambiente seguro, desde la manipulación hasta la autenticación de una copia de un documento que tuvo su complemento digital comprometido.

Palabras-Clave: Archivística; Archivística Computacional; Documento de Archivo Digital.

THE CHAIN OF BITS IN THE LIGHT OF AUTHENTICITY: BINARY CONTENT AS A CENTRAL ELEMENTO OF THE COMPOSITION OF NATO-DIGITAL ARCHIVAL RECORD

This article is dedicated to the analysis of the theoretical foundations of Computer Science regarding the architecture of computers in the scope of transmission of digital archival records and of Archival Science related to the authenticity of digital records. This link between the two areas arises the Computational Archival Science since the end of the 20th century there has been a growth in the adoption and management of digital records. However, the architecture of the native-digital archival records still remains a poorly explored study object, which demonstrates the fundamental relationship with the binary content that is pointed out in the literature as vulnerable to loss of form and content data according to events, and, therefore, a challenge for digital preservation in the Archival Science scope. Therefore, it is worth reflecting on the following question: “what is the relationship of the binary content with the authenticity of the native-digital archive document? This qualitative research was operationalized through bibliographic research of Brazilian and foreign works of Archival Science and Computer Science that sequentially address digital records, authenticity, computer architecture and binary arithmetic. The results suggested that the form of the record is not necessarily linked to authenticity and that, because the binary content is always subject to changes that jeopardize its integrity and, at the same time, is the representative element of the record’s originality that must be preserved in a safe of tampering environment to authenticate record’s copies that had its digital complement compromised.

Keywords: Archival Science; Computational Archival Science; Digital Records.

1 INTRODUCCIÓN

A utilização de tecnologia da informação para a produção de documentos têm sido um desafio sob diversos aspectos na área de informação na contemporaneidade. A multiplicação das possibilidades de representação, organização evoca aos profissionais da área a necessidade de se aproximar dos conceitos, metodologias e teorias deste universo (Ramalho, 2011), de modo a compreender o inter-relacionamento destes com os parâmetros disciplinares próprios.

Diante deste urgente “ajuste tecnológico” (Saracevic, 1996, p.41) ao refletir acerca do conceito de documento de arquivo

digital remete ao redescobrimiento das características do documento de arquivo e, conseqüentemente, alterações em nuances teóricas da literatura arquivística sob o escopo de gestão ou preservação de documentos de arquivo digitais cujas o seu processo de realização requer a avaliação de fronteiras e diálogos entre a Arquivologia, Ciência da Informação e a Ciência da Computação. Logo, diante da transformação digital nos arquivos desde o fim do século XX até os dias atuais, notou-se um vínculo entre as áreas devido ao crescimento do volume de documentos de arquivo digitais que viria a contribuir com o recente surgimento da Arquivologia

Computacional e é neste escopo que este artigo se insere.

Arquivologia Computacional, para Marciano et al. (2018, p. 2), significa um campo transdisciplinar que tem como objeto a aplicação de métodos e recursos computacionais para fins de processamento, análise, armazenamento, preservação a longo prazo e acesso a documentos de arquivo digitais e apoio às funções arquivísticas, como avaliação, arranjo, descrição, além de preservação e decisões de acesso, bem como avaliação dos impactos de todas estas circunstâncias nas funções arquivísticas, no uso organizacional, na preservação da autenticidade dos documentos e na sociedade.

Deste modo, Payne (2019, p. 5) esclarece que a Arquivologia Computacional tem como objetivo a permanência do documento de arquivo no tempo e espaço através da preservação digital, o aspecto humano da informação e da sua interação com a tecnologia por meio de construções sociotécnicas de redes de informação e pesquisa, e o entendimento da teoria da computação e otimização de design de sistemas a partir de práticas de otimização de design e Computação. Como resultado, esta ciência pode encontrar fatos relacionados ao ato da ação que o documento de arquivo registra em favor de averiguar a autenticidade e a confiabilidade; concluir informações acerca da interação humana com uma variedade de configurações de designs de sistemas em favor da otimização da experiência da interação, e avaliar a otimização de viabilidade, estrutura de expressão e mecanização de algoritmos, sistemas e redes em prol da otimização, processamento, aquisição, representação, processamento, armazenamento, e acesso da informação.

Sendo assim, cabe a compreensão do documento de arquivo digital e a sua arquitetura na qual seu principal ator são os bits que o representam em um computador cujo funcionamento é conforme sistema numérico binário como fundamento teórico basilar do

escopo de produção de documentos digitais, portanto, compreendê-lo poderá oferecer fundamentos teóricos a serem utilizados, por exemplo, no aprimoramento de políticas e práticas de preservação digital da Arquivologia Computacional.

De acordo com Rondinelli (2011, pp. 223-226), o conceito de documento arquivístico digital consiste naqueles que são criados conforme atividades realizadas por pessoas físicas e jurídicas em suporte digital acessível por equipamento eletrônico e codificado em bits, de forma oposta ao documento eletrônico que apesar de acessível por instrumento eletrônico, não é codificado em binário. Similarmente à definição do Dicionário Brasileiro de Terminologia Arquivística (Dibrate, 2005, pp. 75), “Documento codificado em dígitos binários, acessível por meio de sistema computacional” e do *Multilingual Archival Terminology* (2012) “Documento codificado em dígitos binários, acessível por meio de sistema computacional”.

Visto isso, há margens para reflexão de que são necessários procedimentos adicionais aos documentos de arquivo digitais, tendo em vista que estes, de acordo com o *International Council of Archives* (ICA) (1997), possuem um nível lógico que contém códigos binários, portanto não são imediatamente evidentes para o usuário humano pois seus dados de forma e conteúdo dependem dos softwares que os decodificam para serem inteligíveis por humanos. Logo, devem ser sempre migrados entre extensões de arquivo em razão da obsolescência tecnológica e, por estarem inseridos em um contexto administrativo, é uma entidade lógica e imaterial que serve como evidência.

Nesse sentido, estes pontos levam a pensar acerca do enraizado vínculo do conteúdo binário do documento de arquivo digital com a preservação digital devido à necessidade de comprovação da autenticidade, peculiaridade esta que pode ser volátil e depender dos cuidados tomados com o documento durante o seu armazenamento e

preservação a longo prazo. Deste modo, a presente pesquisa partiu da seguinte pergunta: “qual a relação do conteúdo binário com a autenticidade do documento de arquivo digital?”. Para respondê-las, traça-se o conceito de documento de arquivo digital e de conteúdo binário e a sua relação fundamental com este

elemento, conceitua-se a relevância do conteúdo binário à luz da ótica da Diplomática e, a partir disso, avalia-se as necessidades tecnológicas e/ou teóricas em favor da autenticidade dos documentos de arquivo digitais.

2 O CONCEITO DE DOCUMENTO DE ARQUIVO DIGITAL E O CONTEÚDO BINÁRIO COMO CARACTERÍSTICA ESPECIAL

De acordo com Rondinelli (2011, p. 223-226), o conceito de documento arquivístico digital consiste naqueles que são criados conforme atividades realizadas por pessoas físicas e jurídicas em suporte digital acessível por equipamento eletrônico e codificado em bits, de forma oposta ao documento eletrônico que apesar de acessível por instrumento eletrônico, não é codificado em binário. Similarmente à definição do Dicionário Brasileiro de Terminologia Arquivística (DIBRATE, 2019, p. 75), “Documento codificado em dígitos binários, acessível por meio de sistema computacional” e do *Multilingual Archival Terminology* (2012) “Documento codificado em dígitos binários, acessível por meio de sistema computacional”.

Por outro lado, para Bellotto e Camargo (2012, p. 40) não há diferenciação entre documento de arquivo eletrônico e documento de arquivo digital, visto que só consta o primeiro termo, sua definição consiste em “Documentação (1) cujo conteúdo, registrado em suportes especiais é acessível apenas por computador”. De maneira semelhante, porém, mais detalhada, o *Dictionary of Archives Terminology* (DAT) (2020, tradução nossa) conceitua apenas documento de arquivo eletrônico, sendo a definição “dados ou informações que foram capturadas ou fixadas para armazenamento e manipulação em um sistema automatizado que requer o uso de um sistema para renderizá-lo de forma inteligível para uma pessoa” e considera o termo documento de arquivo digital como sinônimo deste.

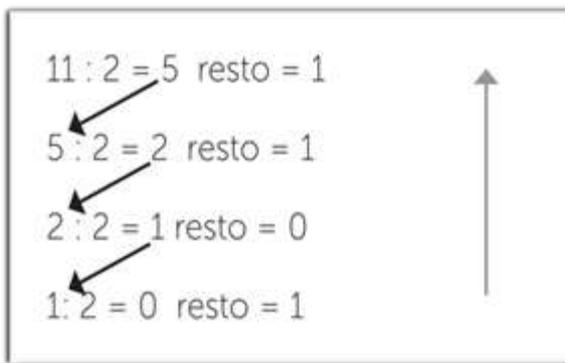
Apesar das definições propostas serem caracterizadas pela diferenciação ou não entre dois termos apresentados, para todas há convergência no ponto de que o documento de arquivo digital é acessível por computador/sistema computacional. Contudo, em nem todas o conteúdo binário é mencionado, embora já seja possível de se deduzir de que a composição do documento de arquivo digital possua este elemento, uma vez que um computador utiliza do sistema numérico binário tanto em seu funcionamento quanto para armazenamento de, por exemplo, o teor e aparência de um documento de arquivo.

Isto é, de acordo com Stallings (2010, pp. 249-250), a aritmética do computador é utilizada no modo como os números são representados em o formato binário, bem como nos algoritmos usados para operações aritméticas básicas, tais quais adição, subtração, multiplicação e divisão tanto para números inteiros quanto números de ponto flutuante, onde neste último caso a maioria dos processadores utilizam o padrão IEEE 754 em um formato de 32 ou 64 bits. Contudo, para fins de armazenamento, como por exemplo, de documentos de arquivo, a representação é realizada apenas a partir de números inteiros (Stallings, 2010, pp. 249-250).

Sendo assim, no sistema numérico binário, quaisquer números do sistema numérico decimal, bem como outros sistemas numéricos, podem ser representados apenas com os dígitos zero e um quando realizada a conversão, no caso de uma cadeia de 8 bits, por

exemplo, podem ser representados números decimais de 0 a 255. Mais explicitadamente, os binários representam um valor, assim como os decimais, portanto, um binário pode possuir o mesmo valor de um decimal, pois trata-se apenas de representações em diferentes sistemas de numeração. Ou seja, o binário é equivalente aos restos das divisões de um decimal, onde o primeiro dígito binário é o último resto encontrado na divisão e o último dígito da cadeia de bits é o primeiro resto da divisão (Stallings, 2010, pp. 250); (Santos & Gibim, 2015, pp. 13) (Figura 1).

Figura 1: Conversão de um valor inteiro da base decimal para base binária.



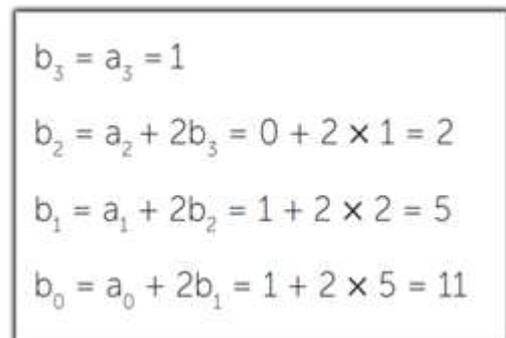
Fonte: Santos e Gibim (2015, p. 13).

Visto que valores inteiros são utilizados para representar cadeias de bits de documentos de arquivo, para fins de explicação, como observado, para realizar a conversão de um valor inteiro da base decimal para a base binária, basta inicialmente realizar uma divisão inteira do valor do dividendo em questão por dois, neste caso está sendo submetido o valor 11 à operação matemática, e após isso dividir o seu quociente e realizar esta última ação novamente e assim sucessivamente até obter o resto da divisão de cada divisão. Com isso, ordena-se os restos das divisões de trás para frente para obter o mesmo valor na base binária. Logo, neste caso, o valor 11 na base decimal significa 1011 na base binária.

Para realizar a conversão do valor da base binária para a decimal, cada posição dos bits representa os coeficientes de potências de dois a serem colocados na operação

matemática, como por exemplo: $(a_j a_{j-1} \dots a_2 a_1 a_0)$. A partir disso, o resultado é obtido através da soma de parciais multiplicações entre cada dígito por potências de dois em que seus expoentes são estas posições (Santos & Gibim, 2015, p. 12) (Figura 2).

Figura 2: Conversão de um valor inteiro da base decimal para base binária.



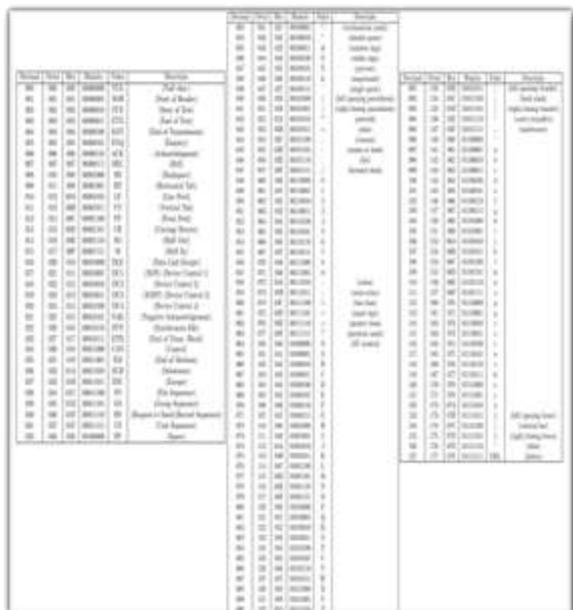
$b_3 = a_3 = 1$
 $b_2 = a_2 + 2b_3 = 0 + 2 \times 1 = 2$
 $b_1 = a_1 + 2b_2 = 1 + 2 \times 2 = 5$
 $b_0 = a_0 + 2b_1 = 1 + 2 \times 5 = 11$

Fonte: Santos e Gibim (2015, p. 12).

De modo a detalhar as operações realizadas acima, a partir do valor inicial 1011, é possível obter os coeficientes a serem utilizados a partir das posições de cada dígito, ou seja, $(1_3 0_2 1_1 1_0)$ que serão utilizados para multiplicar cada um destes dígitos por potências de 2 que contém estes valores encontrados nas posições como expoentes. Logo, a equação deve ser realizada da seguinte forma para obter o valor 11 a partir deste binário: $11 = 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0$.

Quanto aos caracteres textuais do documento de arquivo, para Stallings (2010, p. 293), em razão da dificuldade de armazenamento e processamento desses dados em computadores, são codificados e decodificados em binários de acordo com um código de caracteres onde cada caractere representa um valor inteiro em binário, como por exemplo, o *International Reference Alphabet* (IRA) e o *American Standard Code for Information Interchange* (ASCII), onde cada caractere é representado por um padrão de sete bits, como indicado na Figura 3, e o *Extended Binary Coded Decimal Interchange Code* (EBCDIC), onde cada um significa oito bits.

Figura 3: Tabela da American Standard Code for Information Interchange (ASCII).



Fonte: Bahiana (2009, p. 12-14).

Deste modo, segundo Rondinelli (2011, p. 223) os documentos arquivísticos digitais são codificados na linguagem binária e para tornarem-se acessíveis através da linguagem alfabética, é necessária a intermediação de programas computacionais que são igualmente codificados em bits. Afinal, segundo Stallings (2010, p. 350), programas contém uma sequência de instruções que são convertidas de linguagem de programação para linguagem de montagem que possui mapeamento das instruções e dados percorridos pelo programa para valores em binários que são armazenados em um determinado endereço da memória, tal é o caso da linguagem Assembly (Figura 4).

Figura 4: Simulação de endereços, instruções e conteúdo binário de um programa de linguagem de montagem Assembly.

Endereço	Conteúdo		Endereço	Conteúdo
101	0010	0010	101	2201
102	0001	0010	102	1202
103	0001	0010	103	1203
104	0011	0010	104	3204
201	0000	0000	201	0002
202	0000	0000	202	0003
203	0000	0000	203	0004
204	0000	0000	204	0000

(a) Programa binário

Endereço	Instrução	Operando
101	LOD	201
102	ADD	202
103	ADD	203
104	STA	204
201	DAT	2
202	DAT	3
203	DAT	4
204	DAT	0

(c) Programa simbólico

Rótulo	Operação	Operando
FORMUL	LOD	I
	ADD	J
	ADD	K
	STA	N
I	DATA	2
J	DATA	3
K	DATA	4
N	DATA	0

(d) Programa assembly

Fonte: Stallings, 2010, p. 350.

Abreu (2018, pp. 63) complementa a perspectiva e afirma que na Diplomática contemporânea, o documento digital-original dura apenas o tempo em que foi recebido, lido e salvo pela primeira vez, sendo considerado as próximas reproduções através do software de exibição do documento apenas cópias, uma vez que Durante & MacNeil (1996, p. 63) o definiram como o primeiro documento a ser recebido com a ação registrada identificável. Complementarmente, Santos, Innarelli & Sousa (2007, pp. 28), ressaltam que a originalidade do documento digital é afetada por ser necessário softwares para renderizar um documento e estes podem se tornar obsoletos.

Entretanto, caso os bits do documento sejam devidamente mantidos, a ideia de que o documento deixa de ser original a partir da sua segunda reprodução pelo software pode ser passível de discussão, considerando que o documento de arquivo digital é acessível por software que realiza a decodificação do conteúdo binário para caracteres de padrões de teclado para leitura e codificação de caracteres para binários no momento da escrita, caracteres estes que contém o teor e a aparência do documento que são interpretados a partir do conteúdo binário e, portanto, pode-

se, através disto, provar que o documento se manteve inalterado desde que foi recebido. Contudo, esta arquitetura que demarca a peculiaridade do documento de arquivo digital possui uma série de desafios que requerem complementos a partir de inovações tecnológicas que contornem eventos que

possam ocasionar modificações no conteúdo binário do documento de arquivo e reflexões acerca da teoria arquivística no ambiente digital em favor de aspectos principalmente relacionados à autenticidade do documento de arquivo.

3 O PARADOXO DA AUTENTICIDADE DOS DOCUMENTOS DE ARQUIVO DIGITAIS SOB A ÓTICA DA DIPLOMÁTICA E A VULNERABILIDADE DOS BITS

A qualidade de um arquivo depende da manutenção da autenticidade dos seus documentos por meio da prova de uma linha contínua de custodiantes responsáveis. Ou seja, um documento contém uma forma, é elaborado ou usado para registrar uma ação e é preservado sob própria custódia para servir de evidência para o produtor ou pessoas envolvidas com a ação. Assim, arquivos não são criados para interesses ou informação posteriores. Isto é, o arquivo possui uma série de qualidades: a primeira é a imparcialidade, por serem criados para registro de quaisquer evidências de ações de contextos jurídicos ou administrativos; a segunda é a autenticidade, que depende de uma preservação dos arquivos em custódia oficial, em outras palavras, *unbroken chain* que significa dizer uma custódia ininterrupta do acúmulo natural de documentos realizada por um custodiador confiável, podendo ser o produtor inicial ou, quando transferidos os documentos, uma autoridade administrativa, pessoa ou escritório associados à função legada. Este ponto garante a autenticidade dos documentos porque descarta quaisquer possibilidades de falsificação (Jenkinson, 1937).

Em contrapartida, Cook (2001, p. 19-20) afirma que no caso dos documentos de arquivo digitais, não há como ter confiança na continuidade da cadeia de custódia e para fins de comprovação de autenticidade, deve ser preservada a proveniência ou sua ordem em séries ou sistemas junto aos elementos da forma dos documentos. Elementos estes que devem ser analisados à luz da Diplomática, como sugere Duranti (2015, p. 197), uma antiga

ciência com novos usos a qual, a partir de suas análises, os antigos diplomatas criaram uma metodologia de análise de formas documentais a fim de esclarecer a autenticidade dos documentos porque notou-se que possuem semelhantes formas apesar das diferenças de proveniência e procedência, isto é, esta forma é composta pelos elementos intrínsecos e extrínsecos.

Mais detalhadamente, Tognoli (2014, p. 28-43) define que a Diplomática Clássica se iniciou XVII, na França, quando Jean Mabillon publicou a primeira obra que continha critérios específicos para a crítica textual em documentos, *De re diplomatica* (1681), que continha os conceitos de elementos extrínsecos (as variações de tipos de documentos, de material e de escrita que compõem os arquivos de ordens religiosas visitadas pelo autor) e elementos intrínsecos (subscrições, datas, selos e outros) ao documento, definindo então para esta ciência um caráter prático-jurídico onde analisava-se a autenticidade de documentos de arquivo de valor jurídico, tais como diplomas e contratos.

A partir de então, começa a perder esta característica gerada, sobretudo, por disputas de propriedades por eclesiásticos e imperadores entre os séculos XVIII e XIX através da Revolução Francesa a qual houve um fomento à pesquisa documental com a abertura dos arquivos para aos pesquisadores, inicialmente na França, a fim de que formassem uma identidade nacional por meio dos estudos históricos, além de maneira aproximadamente paralela, na Alemanha, ocorre o surgimento da

palavra *Urkunde* (testemunho escrito, redigido segundo uma forma determinada) como objeto da Diplomática na literatura europeia através das publicações de Theodor Von Sickel e Julius Ficker a qual marcou a transição para a Diplomática Moderna por meio da sofisticação dos estudos de análise diplomática. Posteriormente, os estudos diplomáticos foram ampliados por Georges Tessier, na França em 1930, ao considerar o novo termo *actes instrumentaires* que engloba documentos preparatórios ou redigidos para fins de prova jurídica autênticos ou de autenticação privada que precedem um ato jurídico de uma entidade pública ou privada e, complementarmente, Robert-Henri Bautier, em 1961, propõe que todos os documentos de arquivo e não somente os de valor histórico deveriam ser considerados objetos da diplomática (TOGNOLI, 2014, p. 50-100).

Para terminar, a evolução para a Diplomática Contemporânea ocorreu por meio de Paola Carucci e Luciana Duranti, na Itália, em 1987, ao expandir as ideias de Bautier ao considerar também os documentos que não possuíam valor jurídico como objetos da Diplomática por meio do livro denominado *Il documento contemporaneo: Diplomatica e criteri di edizione* (1987). Sequencialmente, em 1989, no Canadá, Luciana Duranti dissemina o estudo através de sua obra *Diplomatics: new uses for an old science*, primeira publicação a qual demonstra a aplicação do método aos documentos norte-americanos e enfatiza que a análise diplomática deve ser aplicada em somente documentos de arquivo, ou seja, criados a partir de uma atividade e deve ser adaptada aos princípios dos manuais clássicos e aos documentos contemporâneos a fim de elucidar o contexto de produção que o documento a partir do método (Tognoli, 2014, p. 106-111).

Sob a nuance da evolução da Diplomática Contemporânea a partir da perspectiva de Duranti (2009, p. 2528), o documento de arquivo digital deve possuir as seguintes características: (a) forma fixa; (b)

conteúdo estável; (c) relação orgânica entre documentos de arquivo armazenados ou não em sistemas; (d) contexto de criação identificável; (e) envolvimento de cinco pessoas, tais quais, autor, redator, destinatário, originador e produtor; (f) ação que o documento registra, como por exemplo, uma tomada de decisão ou suporte às ações registradas por outros documentos.

Nesse sentido, dentre as características identificáveis através de uma análise diplomática de um documento de arquivo digital, enfatiza-se a forma fixa e o conteúdo estável como centros da discussão. Porque se considera um documento de arquivo de forma fixa aquele que seu conteúdo binário for capaz de reproduzir a mesma apresentação do documento de arquivo de quando foi salvo, mesmo que houvesse modificação de extensão do arquivo ou quando o documento pode representar o mesmo conteúdo com apresentações pré-determinadas, como gráficos e tabelas, portanto, mesma cadeia de bits para todos os tipos de apresentação (Duranti, 2010, p. 11).

Para fundamentar esta discussão, vale lembrar que Heredia Herrera (2017) compara as divergências entre as definições de forma fixa as quais algumas delas podem possuir conflitos conceituais entre o documento de arquivo analógico e o documento de arquivo digital. O conceito de forma, atribuído pelos diplomatas, significa o estado de transmissão, ou seja, cópia, minuta, rascunho, original e outros; enquanto para os arquivistas consiste em um conjunto de fórmulas ou cláusulas responsáveis por distinguir o tipo documental; quanto à concepção para normas internacionais, forma é uma maneira de distinguir um campo de descrição; por fim, para a autora, define-se como um padrão de acumulação de documentos, como processo, dossiê e outros.

De igual forma à definição demarcada pelos diplomatas a qual menciona-se um estado de transmissão que remete à aparência do documento, Bellotto e Camargo (2012, p. 50)

definem o atributo como “Estágio de preparação e de transmissão de documentos”, bem como o *Multilingual Archival Terminology* (2012), “Característica de um documento arquivístico que assegura que sua aparência ou apresentação documental permanece a mesma cada vez que o documento é manifestado, ou pode ser alterada segundo regras fixas”. Assim como DAT (2020, tradução nossa) nos termos equivalentes de língua inglesa “transmission form” e “physical form”, “A aparência geral, configuração ou formato de um documento, independentemente de seu conteúdo intelectual, quando recebido (após a transmissão)”, embora neste caso haja a diferenciação da aparência ao teor documental através do termo “intellectual form” que significa “A soma dos atributos formais de um documento que representam e comunicam os elementos da ação em que o documento está envolvido e de seu contexto imediato, tanto documental quanto administrativo”.

Contudo, externamente à perspectiva de forma dos dicionários de terminologia arquivística, nota-se algumas divergências em relação às definições supracitadas onde algumas delas se aproximam das definições de formato das publicações indicadas no Quadro 1.

Quadro 1: Definições de forma e formato

Autor	Definição de forma e formato
Vicenta Cortés Alonso (1989)	Forma: tradição documental Formato: configuração física do documento que depende da matéria, suporte e da manipulação para criar (o documento)
ISAD (G) (2000)	Forma: (form) como característica física (Aquarela, audiovisual e etc.) – sinônimo de tipo documental (não obrigatório)
López e Gallego (2007)	Forma: estruturação do conteúdo (tipo documental) e tradição documental (característica externa) Formato: característica física
José Ramón Cruz Mundet (2011)	Forma: atributo interno próximo ao tipo documental e tradição documental Formato: característica física
Antonia Heredia (2011)	Forma: próximo ao tipo documental e também próxima à tradição documental Formato: características físicas. Enxerga correspondência com o elemento de descrição “volume e suporte” da ISAD (G)
Alicia Barnard e Manuel Voutssas (2014)	Forma do documento de arquivo: especificações das formas documentais para os distintos tipos de documentos de um produtor. Regras de representação que estabelecem a aparência de uma entidade documental (Interno e externo) Formato: -
Isabel Seco Campos (2017)	Forma: tradição documental Formato: configuração física imprescindível no contexto digital
Lluis-Esteve Casellas (2017)	Forma documental: tipo de acumulação de documentos (processos ou dossiês) Formato: -
López e Gallego (2007) e Herrera (2011)	Forma documental/forma básica: maneira de reconhecer os tipos documentais nas unidades documentais compostas

Fonte: Elaboração própria a partir de Herrera (2017).

Desta maneira, nota-se que a definição de forma, anteriormente definida pelos diplomatas tornou-se, ao longo do tempo, a definição de formato, como ocorre nas definições de Campos (2017), López e Gallego (2007), Herrera (2011) e Cortés Alonso (1989), enquanto foi atribuído um outro sentido à forma no segmento de uma tradição documental relacionada à configuração do tipo documental de acordo com a sua produção,

como definem Cortés Alonso (1989), López e Gallego (2007), Heredia Herrera (2011), Cruz Mundet (2011), Barnard e Voutssas (2014), Campos (2017) e Casellas (2017), o que seria equivalente à definição de “intellectual form” do DAT (2020) exposta anteriormente.

Tendo em vista tais constatações, este padrão para definição de forma definida pela Diplomática pode ser um tanto questionável. Porque além da apresentação não garantir a autenticidade e a proveniência do documento de arquivo, no caso dos documentos de arquivo digitais nem sempre é possível a manutenção da aparência do documento, tal é o caso dos documentos dinâmicos que de acordo com Duranti e Preston (2008, p. 786) significa dizer um documento com conteúdo dependente de dados instantâneos armazenados em bancos de dados e/ou planilhas pelos quais o documento importa estes dados. Em contrapartida à definição de documento de arquivo digital de forma fixa de Duranti (2010, p. 11), onde a flexibilidade acerca da aparência vai até diferentes tipos de aparência pré-definidos em um mesmo documento, o que não ocorre nos documentos dinâmicos.

Desta forma, diante da necessidade de adaptar as antigas características identificadas pela Diplomática que permeiam a autenticidade de um documento de arquivo digital, remete a possibilidade de novos paradigmas acerca da preservação da aparência que depende precisamente da preservação do conteúdo binário do documento de arquivo. Afinal, segundo Santos e Flores (2015, p. 51), quando é realizada, por exemplo, a migração dos

4 METODOLOGIA

A codificação do teor do documento de arquivo leva a pensar acerca do enraizado vínculo do conteúdo binário do documento de arquivo digital com a preservação digital devido à necessidade de comprovação da autenticidade, peculiaridade esta que pode ser volátil e depender dos cuidados tomados com o documento durante o seu armazenamento e

documentos de arquivo entre formatos devido a eventuais obsolescências, isto é, a extensão que define por quais softwares o documento em questão pode ser acessado, ocorre alterações nas sequências de bits que podem gerar perdas de apresentação e do teor do documento. O que pode tornar a fidedignidade do documento de arquivo questionável e, conseqüentemente, dependente de outros tipos de estratégia de preservação digital, como a emulação e o encapsulamento que embora preservem a cadeia de bits, dependem de softwares e hardwares específicos que podem igualmente tornarem-se obsoletos.

Logo, tanto o descompromisso com a preservação do conteúdo binário quanto a obsolescência de formatos e softwares de leitura de documento podem ser consideradas ameaças à autenticidade dos documentos de arquivo digital, uma vez que estes riscos podem infringir os conceitos tanto de forma quanto de formato definidos pela Arquivologia que estão intimamente relacionados com a autenticidade de um documento de arquivo porque tanto o conteúdo quanto a aparência do documento podem ser afetados. Sendo assim, estes pontos requerem novas alternativas que, simultaneamente, sejam asseguradas a preservação das cadeias de bits no momento de criação do documento de arquivo junto a sua contínua acessibilidade de modo a evitar dependência de softwares e hardwares específicos para que não seja preciso adotar um meio que seja necessário um procedimento que modifica a sequência de bits, como a migração.

preservação a longo prazo. Deste modo, a presente pesquisa partiu da seguinte pergunta: “qual a relação do conteúdo binário com a autenticidade do documento de arquivo digital?”. Para respondê-las, traça-se o conceito de documento de arquivo digital e de conteúdo binário e a sua relação fundamental com este elemento, conceitua-se a relevância do

conteúdo binário à luz da ótica da Diplomática e, a partir disso, avalia-se as necessidades tecnológicas e/ou teóricas em favor da autenticidade dos documentos de arquivo digitais.

5 RESULTADOS

Elementos atrelados à autenticidade do documento de arquivo digital podem ser mantidos através da preservação dos bits, tal é o caso da proveniência que depende da manutenção da autenticidade, confiabilidade, história de criação, propriedade, acessos. Contudo, a preservação do conteúdo binário é um desafio por ser um componente que se degrada conforme algumas ações e, em contrapartida, é necessário preservar estes atributos para comprovar que os documentos de arquivo são autênticos, incorruptos e de integridade mantida, o que os binários seriam capazes de provar por conterem os dados exatos do momento de criação. Logo, não há como assegurar que softwares de armazenamento de arquivos preservem com precisão este elemento (Lemieux, 2016, p. 84).

De acordo com Tognoli (2014, p. 43), o princípio da proveniência tem berço na Europa de forma simultânea à teoria da Diplomática. No Antigo Regime, inclui-se este princípio através da consideração da instituição produtora como elemento crucial do documento de arquivo individual a ter a sua autenticidade analisada conforme a teoria da Diplomática. Logo, perder, por exemplo, bits de representação do protocolo inicial que contém o timbre do órgão emissor significaria a perda da autenticidade do documento.

Ao partir da garantia do princípio da proveniência, obtém-se o a confiabilidade porque este elemento significa o conhecimento da origem de algo ou alguém e é por isso que é

6 CONSIDERAÇÕES FINAIS

No contexto de uma *blockchain*, segundo Mittelbach & Fischlin (2021, p.5) a

Portanto, a metodologia desta pesquisa consistiu em natureza qualitativa e operacionalização através de pesquisa bibliográfica por meio das bases de dados Google Acadêmico, Taylor & Francis, SCOPUS e *Scientific Electronic Library Online* (SciELO).

um conceito próximo à proveniência. Com isso, no âmbito da Ciência da Computação houve o crescimento de desenvolvimento de aplicações de rastreamento e análise de proveniência de dados em redes distribuídas baseadas em *blockchain* por ser uma tecnologia capaz de promover a confiabilidade, acurácia e autenticidade dos registros (Lemieux, 2021, p. 111).

A tecnologia *blockchain*, quando aplicada a documentos de arquivo, recebe dados de entrada relacionadas ao documento e a proveniência destes dados passa a ser de vários sistemas que estiverem sendo executados em cada computador que compor a rede esta pode ser utilizada para concordar que um dado a ser inserido na rede é exato (Lemieux, 2021, p. 116).

Portanto, uma vez que a *blockchain* pode salvar o seu conteúdo binário do momento de criação do documento entre os metadados dos *hashs* armazenados nos blocos, segundo Lemieux (2021, p. 116), a tecnologia pode ser utilizada para autenticação e comparação dos documentos e para verificar originalidade do documento. Porque além de gerar os *hashs* que significa a criptografia destes dados, utiliza-se de um mecanismo de consenso com políticas de tomada de decisão entre os participantes da rede que pode ser executado em sistemas automatizados, manuais ou híbridos, bem como promover transparência, imutabilidade e rastreabilidade destes registros através da constante verificação e confirmação dos blocos.

função *hash* tem o papel de armazenamento seguro de dados para garantia da sua

integridade por meio de criptografia também com base em esquemas de assinatura digital assimétrica em *blockchain* cujo sistema é usado como fonte de entropia para a criptografia do código *hash*.

Isto é, os dados acerca do remetente e destinatário também podem ser armazenados, assim como outros dados referentes a ações que ocorrem de forma entrópica entre os participantes da rede a depender da escolha do algoritmo criptográfico, como o exclusivo conteúdo binário de um documento, no âmbito dos arquivos.

Deste modo, armazenar o conteúdo binário através de *hashs* armazenados em uma

blockchain pode ser vantajoso tanto no sentido de autenticação de documentos de arquivo considerados cópias por terem tido sua cadeia de bits adulterada em razão de algum evento, como por exemplo, a migração entre diferentes formatos quanto para recuperação de documentos de arquivo eliminados indevidamente, uma vez que o *hash* da *blockchain* armazenará o conteúdo binário de forma inalterável assegurando a confiabilidade e autenticidade dos dados enquanto a rede tiver uma população de contribuidores com o propósito de preservação de documento a longo prazo para validação e confirmação dos dados a serem inseridos nos blocos da rede.

REFERÊNCIAS

- Abreu, J. P. L. (2018). *Existir em bits: arquivos pessoais nato-digitais e seus desafios à teoria arquivística*. São Paulo: Associação dos Arquivistas de São Paulo.
- Bahiana, M. (2009). *Introdução à Computação Científica em C*. Rio de Janeiro: Universidade Federal do Rio de Janeiro.
- Bein, B. (2006). Entropy. *Best Practice & Research Clinical Anaesthesiology*, 20 (1).
- Bellotto, H. L & Camargo, A. M. A. (2012) *Dicionário de Terminologia Arquivística*. (3.ed.) São Paulo: Associação dos Arquivistas Brasileiros.
- Cook, T. (2001). Archival science and postmodernism: new formulations for old concepts. *Archival Science*, 1, pp.3-24.
- Dictionary of Archives Terminology, s.v. Recuperado em 21-10-2022, <https://dictionary.archivists.org/index.html>.
- Duranti, L. (2015). Diplomática: novos usos para uma antiga ciência. *Acervo*, 28(1), pp.196-215.
- Duranti, L. (2009). *Encyclopedia of Library and Information Sciences* (4ª ed.) Londres, Taylor & Francis Group.
- Duranti, L. & Preston, R. (2008). *International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records*. Padova: Associazione Nazionale Archivistica Italiana.
- Duranti, L. (2010). *Structural and formal analysis: the contribution of diplomatics to archival appraisal in the digital environment*. In Hill, J. *The Future of Archives and Recordkeeping: A Reader*. (pp. 65-68). Londres: Facet.
- Duranti, L. & Macneil, H. (1996). The Protection of the integrity of electronic records: an overview of the UBC-MAS research Project. *Archivaria: ACA*, 42, pp.46-67.
- Heredia Herrera, A. (2017). Forma y formato de los documentos de archivo. *Boletín ANABAD*, LXVII (4), pp. 39-51. Recuperado em 2022-08-01, de <https://dialnet.unirioja.es/servlet/articulo?codigo=6449987>

- International Council of Archives (1997). Guide For Managing Electronic Records From an Archival Perspective.
- Institute of Electrical and Electronics Engineers (2020). International Standard - Floating-point arithmetic.
- Jenkinson, H. (1937). *A Manual of Archive Administration*. (2^a ed.) Londres: Percy Lund, Humphries & Co.
- Lemieux, V. (2016) *Building Trust in Information Perspectives on the Frontiers of Provenance*. Vancouver: Springer.
- Lemieux, V. (2021) *Building Decentralized Trust*. Vancouver: Springer.
- Marciano, R. (2018). Archival records and training in the age of Big Data. *Perspectives on the Future of Library and Information Science Education*, 44b, pp.179-199.
- Mittelbach, A. & Fischlin, M. (2021). *The Theory of Hash Functions and Random Oracles*. Darmstadt: Springer.
- Multilingual Archival Terminology. International Council on Archives (2012). Recuperado em 21-10-2020, de <http://www.ciscra.org/mat/mat>.
- PAYNE, N. (2018). Stirring The Cauldron: Redefining Computational Archival Science (CAS) For The Big Data Domain. *IEEE International Conference On Big Data*.
- Rondinelli, R. C. (2011) *O Conceito de documento arquivístico frente à realidade digital: uma revisitação necessária*. Tese de doutorado, Instituto Brasileiro em Ciência e Tecnologia, Universidade Federal Fluminense, Niterói.
- Santos, B. V., Inarelli, H. C. & Sousa, R. T. (2007). B. *Arquivística: temas contemporâneos*. Distrito Federal: Senac.
- Santos, J. C., Gibim, G. F. B (2015). *Cálculo Numérico*. Londrina: Editora e Distribuidora Educacional S.A.
- Stallings, W. (2010). *Arquitetura e organização de computadores*. (8.ed.) São Paulo: Pearson Pratices Hall.
- Tognoli, N. B. A. (2014). *Construção teórica da Diplomática: em busca da sistematização de seus marcos teóricos como subsídio aos estudos arquivísticos*. São Paulo: Editora UNESP.
- Ramalho, R. A. S. (2011) Tecnología y sociedad: responsabilidad y retos de los profesionales de la información. *Revista EDICIC*, 1 (2), pp.122-135. Recuperado em 2022-08-01, de <https://ojs.edicic.org/index.php/revistaedicic/article/view/41>.
- SARACEVIC, T. (1996). Ciência da informação: origem, evolução e relações. *Perspectivas em ciência da informação*, 1(1), p.41-62.