Versão

DIPLOMÁTICA DE DOCUMENTOS NATO DIGITAIS: A CONSIDERAÇÃO DA FORMA DOCUMENTAL NO AMBIENTE DIGITAL *

Corinne Rogers I Escola de Biblioteconomia, Arquivologia e Estudos da Informação (SLAIS) da University of British Columbia, Vancouver, Canadá

Tradução: Juan Bernardo Montoya Mogollón

Revisão: Márcia Beatriz Carneiro Aragão

RESUMO

Propósito – este artigo objetiva explorar um novo modelo de "documento de arquivo" analisando seus atributos a partir de uma análise técnica de documentos de arquivos no formato digital. O estudo compara as caraterísticas centrais necessárias para nomear um objeto digital como "documento de arquivo" em relação com a diplomática, ou de "prova" quanto à análise forense digital. Este estudo divide os documentos digitais em três camadas de abstração, a saber: conceitual, lógica e física. Nossa proposta é aplicar a diplomática de documentos de arquivo digitais, para identificar os principais elementos em cada um desses níveis de abstração.

Desenho/metodologia/abordagem – A diplomática digital é resultado do projeto da Pesquisa Internacional sobre Documentos de Arquivo Autênticos e Permanentes em Sistemas Eletrônicos (InterPARES)[1] , a qual fornece para os arquivistas uma metodologia que analisa a identidade e a integridade dos documentos de arquivos digitais em sistemas eletrônicos, como também avalia sua autenticidade (Duranti e Preston, 2008; Duranti, 2005) e rastreia sua proveniência.

Descobertas– os documentos de arquivo digitais estão estruturados em: dados (conteúdo) gerados pelo usuário, metadados gerados pelo sistema que identificam a fonte e a localização, metadados gerados por aplicativos que gerenciam a aparência e o desempenho do documento de arquivo (como, por exemplo, o formato de arquivo nato digital), metadados gerados por aplicativos que descrevem os dados (como por exemplo os metadados do sistema operacional utilizado pelo arquivo), e metadados que descrevem os dados gerados pelo usuário. A diplomática digital, baseada nos princípios da diplomática tradicional, pode subsidiar na identificação de documentos de arquivos digitais por meio de seus metadados e determinar quais deles são necessários para serem capturados, gerenciados e preservados.

Originalidade/valor- O valor e a originalidade deste artigo estão na aplicação dos princípios diplomáticos para uma visão técnica desconstruída dos documentos de arquivo digitais, por meio dos metadados funcionais que avaliem a identidade e autenticidade desses documentos arquivísticos digitais.

Palavras chaves - Metadados, Gestão documental, Diplomática, Forense Digital.

Tipo de artigo – Artigo conceitual

* Artigo traduzido do original: Corinne Rogers, (2015),"Diplomatics of born digital documents – considering documentary form in a digital environment", Records Management Journal, Vol. 25 Iss 1 pp. 6 – 20 Permanent link to this document: http://dx.doi.org/10.1108/RMJ-03-2014-0021

[1] *International Research on Permanent Authentic Records in Electronic Systems (InterPARES)* [N.T.].

ABSTRACT

Purpose – This paper aims to explore a new model of "record" that maps traditional attributes of a record onto a technical decomposition of digital records. It compares the core characteristics necessary to call a digital object a "record" in terms of diplomatics or "evidence" in terms of digital forensics. It then isolates three layers of abstraction: the conceptual, the logical and the physical. By identifying the essential elements of a record at each layer of abstraction, a diplomatics of digital records can be proposed.

Design/methodology/approach – Digital diplomatics, a research outcome of the International Research on Permanent Authentic Records in Electronic Systems (InterPARES) project, gives archivists a methodology for analyzing the identity and integrity of digital records in electronic systems and thereby assessing their authenticity (Duranti and Preston, 2008; Duranti, 2005) and tracing their provenance.

Findings – Digital records consist of user-generated data (content), system-generated metadata identifying source and location, application-generated metadata managing the look and performance of the record (e.g., native file format), application-generated metadata describing the data (e.g., file system metadata OS), and user-generated metadata describing the data. Digital diplomatics, based on a foundation of traditional diplomatic principles, can help identify digital records through their metadata and determine what metadata needs to be captured, managed and preserved.

Originality/value – The value and originality of this paper is in the application of diplomatic principles to a deconstructed, technical view of digital records through functional metadata for assessing the identity and authenticity of digital records.

Keywords - Metadata, Records management, Diplomatics, Digital forensics Paper type Conceptual paper

Introdução

A diplomática é a ciência de análise de documentos baseada no estudo sistemático de elementos intrínsecos e extrínsecos da forma documental. É definida como a análise da criação, formas e estados de transmissão de documentos de arquivos, ou records[2] , e sua relação com os fatos representados neles e em relação com seu criador para identificar, avaliar e comunicar a sua verdadeira natureza (Duranti, 1998). A necessidade de um método rigoroso de análise de documentos antigos, se deu pela proliferação de falsificações na Baixa Idade Média; no entanto, não existiu tal método até o século XVII, quando foi estabelecida a fundação da moderna disciplina da crítica diplomática com o trabalho de Dom Jean Mabillon e a publicação de De Re Diplomatica, em 1681. No começo, a disciplina era aplicada unicamente aos documentos probatórios de atos jurídicos, e no século XX, a aplicação da diplomática foi estendida para documentos não jurídicos e dossiês de documentos relacionados com fatos ou atos (Boyle, 1992). O uso da diplomática para avaliar a autenticidade de escritos cada vez mais especializados e a variedade crescente das formas documentais têm levado, sem surpresa, à aplicação de princípios diplomáticos no ambiente digital e ao desenvolvimento de um campo especializado da "diplomática digital". As pesquisas sobre a diplomática digital têm sido desenvolvidas mais extensivamente na Universidade da Colúmbia Britânica (UBC)[4], por meio dos projetos da Pesquisa Internacional sobre Documentos de Arquivo Autênticos e Permanentes em Sistemas Eletrônicos (InterPARES) (Duranti e MacNeil, 1997; Duranti, 2005; Duranti e Preston, 2008).

A ciência diplomática continua com sua relevância até hoje e no contexto digital vem se desenvolvendo em duas direções. A primeira delas tem a ver com a digitalização de fontes históricas e o uso de ferramentas digitais que auxiliem em sua crítica diplomática, por meio da análise de seus elementos formais extrínsecos e intrínsecos (metadados, em terminologia moderna). Estes incluem, por exemplo, ferramentas como o reconhecimento ótico de caracteres para analisar e interpretar fontes históricas, o uso da analítica da visualização para auxiliar na comparação e na análise de grande número de itens, e na produção e na metodologia de edições críticas. Nesse sentido, a diplomática digital é a aplicação da diplomática clássica de documentos tradicionais feita em documentos digitais, e seu avanço teórico foi possível devido às possibilidades da tecnologia digital.

Uma segunda direção do estudo da "diplomática digital" se refere à aplicação da teoria e dos princípios da diplomática tradicional na análise de gêneros de comunicação nato digitais. O projeto InterPARES tem pesquisado requisitos da preservação da autenticidade de documentos de arquivo, criados e/ou mantidos em bases de dados e no gerenciamento de sistemas no decorrer de atividades administrativas (Etapa 1 – 1999-2001), e nos documentos dinâmicos e interativos, produzidos em ambientes digitais no decorrer de atividades artísticas, científicas e governamentais (Etapa 2 – 2002-2007). Os resultados dessas duas etapas foram aplicados em arquivos e em documentos de arquivo, dentro de organizações com limitações financeiras e/ou de recursos humanos, com o objetivo de implementar uma gestão documental bem fundamentada e programas de preservação (Etapa 3 – 2008-2012). O InterPARES tem aprimorado e ampliado a teoria da diplomática tradicional para desenvolver uma ontologia de um documento de arquivo digital, além de ter moldado as atividades necessárias para estabelecer e proteger a autenticidade dos documentos de arquivo ao longo de seu ciclo de vida (o modelo da Cadeia de Preservação) (Duranti e Preston, 2008; Duranti e Thibodeau, 2006). O trabalho continua em 2013 com o lançamento do InterPARES Trust (2013-2018), que estuda as questões da confiabilidade dos documentos de arquivo conservados e usados em ambientes *online*.

Essas diretrizes de diplomática digital são um resultado da pesquisa do InterPARES, o qual fornece aos arquivistas uma metodologia na análise da identidade e da integridade dos documentos de arquivo digitais, em sistemas eletrônicos e tornando, assim, possível avaliar sua autenticidade (Duranti e Preston, 2008; Duranti, 2005) e rastrear sua proveniência. A diplomática digital é apropriada, idealmente, à análise da autenticidade dos documentos de arquivo digitais assim definidos pela arquivologia, porém é limitada quando o assunto da análise é ampliado para a inclusão de objetos digitais que talvez não satisfaça essa definição precisa, mais estrita (Duranti e Endicott-Popovsky, 2010, p.2; MacNeil e Gilliland-Swetland, 2005, p.52).

Um documento de arquivo pode ser definido como um documento – isto é, informação registrada – produzido ou recebido no exercício da atividade prática, como um instrumento ou subproduto dessa

[2] *definição Anglo-Saxônica* [N.T.].
[3] *University of British Columbia (UBC)* [N.T.].

atividade e retido para outra ação ou referência (InterPARES 2-Glossário da palavra "documento de arquivo"). A partir dessa definição, há certas premissas fundamentais sobre os documentos de arquivo. É considerado um documento tradicional ou em papel – isto é, um objeto físico com forma fixa e conteúdo estável. É produto de uma atividade humana ou administrativa e existe em uma estrutura contextual hierárquica e em relação com outros documentos de arquivo. Além disso, sua autoria e pessoas responsáveis podem ser identificadas. Seus elementos intrínsecos e extrínsecos, da forma como estão conectados diretamente com seu meio, e o documento de arquivo original têm valor, pois ele é o primeiro a ser produzido, ou seja, com características de originalidade, ou primitividade[4] , a sua completude e a sua habilidade para atingir seu propósito.

É comum considerar documentos de arquivo, documentos em geral e informações que produzimos e disseminamos na internet, como equivalentes ou similares às formas documentais no mundo físico. A premissa dessa equivalência funcional entre documentos e dados digitais e analógicos, e a autenticidade e fidedignidade dessas novas criações digitais, são muitas vezes julgadas com as mesmas regras. A prática da diplomática é investigativa por natureza. No processo da crítica diplomática, o diplomatista desconstrói um documento para identificar e localizar os elementos que revelam sua proveniência, suas relações, sua confiabilidade e sua autenticidade. Se concordarmos que o núcleo da diplomática é o documento de arquivo e as bases da moderna diplomática são os contextos nos quais o documento de arquivo existe, o ato ou transação do qual ele participa, as pessoas (ou atores não humanos) que participam de sua produção, os procedimentos e as formas documentais que governam sua produção e as relações que o conectam com outros documentos de arquivo, logo podemos entender a análise diplomática como um processo de abstração e de sistematização.

O uso do termo "documento de arquivo" digital pode inicialmente ser mal interpretado, porque o termo está carregado de muitos significados na tradição analógica. No entanto, é possível se referir ao "documento de arquivo digital" nato digital, em que o termo "documento de arquivo" é desconstruído em seus atributos essenciais que transcendem o meio digital. Como acontece com os documentos de arquivo em suporte papel, no ambiente digital os atributos contêm forma fixa, conteúdo estável, vínculos explícitos com outros documentos de arquivo, contextos identificáveis, identificação de pessoas envolvidas na produção do documento e a ação na qual o documento participa. Apesar disso, eles se encontram armazenados e instanciados digitalmente, de maneira que se diferenciam dos suportes analógicos tradicionais, desafiando-os. Como podemos rastrear esses atributos essenciais do "documento de arquivo" desde nosso passado analógico para nosso presente digital?

O mundo digital questiona muitas dessas premissas. A forma de um objeto digital pode não ser fixa nem estável no sentido tradicional. Ele pode ser resultado de uma atividade humana ou administrativa ou como parte de um processo de uma máquina. Os objetos digitais podem ser infinitamente reproduzidos e seu significado e determinação da sua confiabilidade e autenticidade dependem do conhecimento do seu contexto e proveniência. Eles são criados em uma rede horizontal fluida em que a autoria e/ou propriedade são difíceis ou impossíveis de identificar. Os elementos intrínsecos e extrínsecos da forma não dependem ou não estão diretamente ligados ao meio físico; na verdade eles podem não estar imediatamente visíveis. O que costumávamos visualizar no documento, agora está provavelmente oculto. Assim, o próprio documento de arquivo/objeto é somente entendível com a mediação de um ***hardware*** e um ***software.*** As empresas estão se movendo cada vez mais em direção à descentralização de controle de seus ativos digitais, com a adoção da computação na nuvem – os sistemas dos documentos de arquivo eletrônicos têm múltiplos propósitos e altamente interconectados, e o conjunto diverso de dados textuais, visuais e audíveis, que podem estar ou não estar em conformidade com a definição de "documento de arquivo"; ainda precisa ser conservado e preservado e sua confiabilidade e autenticidade devem permanecer protegidas.

Gerenciar e entender os recursos digitais e os sistemas de informação complexos exigem o envolvimento de profissionais em tecnologia da informação. Quando os documentos de arquivo precisam ser analisados em sistemas digitais, é o especialista em forense digital o primeiro a ser chamado, ao invés do arquivista ou do diplomatista. No entanto, em um nível básico, os profissionais em diplomática digital, como em forense digital estão comprometidos em descobrir, entender, descrever e apresentar informações inscritas em meios digitais.

[4] ***primitiveness***, no original [N.T.].

Tanto a forense como a diplomática digital são ciências investigativas. Os investigadores em forense digital procuram, em um disco rígido ou em outro dispositivo de armazenamento, unidades mensuráveis de informação que têm, ou podem ter, valor probatório na reconstrução de eventos. A forense digital é definida como a "aplicação da ciência e da engenharia em questões legais relacionadas com a prova digital" (Pollitt, 2009), e, ainda, o uso de métodos derivados da ciência e comprovados por ela na preservação, coleta, validação, identificação, análise, interpretação, documentação e apresentação da prova digital" (Palmer, 2001). Quais propriedades são necessárias e/ou suficientes para que os materiais digitais sejam viáveis em um contexto investigativo específico? A base da forense digital é a fidedignidade da integridade dos dados, a autenticação, a reprodutibilidade, a não-interferência e a relevância (Mocas, 2004).

A diplomática e a forense digital têm muitas similaridades teóricas e metodológicas e compartilham desafios importantes. Um deles é avaliar a autenticidade e estabelecer o contexto, a proveniência, as relações e o significado dos documentos de arquivo digitais (diplomática), ou provas digitais (forense digital). Reunindo os princípios da diplomática e as ferramentas da forense digital, podemos abordar os problemas da identificação e da análise de documentos de arquivo em ambientes digitais.

Os vínculos iniciais entre os conceitos da forense e as funções da arquivologia foram estabelecidos há quase vinte anos. Em 1994, a autora Elizabeth Diamond relacionou o arquivista com o especialista forense. A autora escrevia que o arquivista, assim como o especialista forense, é uma testemunha especialista, que tem a capacidade de atestar a natureza dos documentos eletrônicos. O arquivista deve ser capaz de "traduzir" esses documentos eletrônicos armazenados em fluxos fragmentados de bits, e ser capaz de atestar sua integridade, da mesma forma com que o cientista forense atesta a integridade dos materiais recolhidos na sua investigação (Diamond, 1994, p.142). No campo da gestão documental, o autor Alastair Irons traçou paralelos explícitos com a forense digital na sua análise dos princípios da forense computacional no contexto das caraterísticas documentais de autenticidade, confiabilidade, integridade e usabilidade. A "Forense computacional", expressa Irons, "deve basear-se nas características de bons documentos de arquivo, níveis e natureza do acesso e a indicação da sua completude[1]". As técnicas da forense digital podem ser usadas para monitorar a integridade, a autenticidade, a confiabilidade e a completude dos documentos de arquivo. Irons também propõe que a forense computacional poderia se beneficiar por meio da aplicação dos modelos teóricos dos documentos de arquivo (Irons, 2006). O projeto da Forense Digital de Documentos de Arquivo (DRF)[5] da Universidade da Colúmbia Britânica (UBC) foi uma colaboração de três anos (abril de 2008-abril de 2011), entre a Escola de Biblioteconomia, Arquivologia e Estudos da Informação (SLAIS)[6] da UBC, a Faculdade de Direito da UBC[7] e a Divisão da Forense Computacional do Departamento de Polícia de Vancouver[8] . A pesquisa abordou os seguintes desafios apresentados pela tecnologia digital para a gestão documental, a arquivística e as profissões ligadas ao Direito: a identificação de documentos de arquivo em sistemas digitais complexos e a determinação da sua autenticidade. A pesquisa foi baseada na metodologia da forense digital e da diplomática arquivística e nos princípios do direito da prova, resultando na proposta de uma nova disciplina: a Forense Digital de Documentos de Arquivo - DRF (Duranti e Endicott-Popovsky, 2010; Duranti, 2009, 2010).

A união interdisciplinar entre a forense digital e a arquivologia encontra-se, neste momento, indo da teoria para a práxis em arquivos e instituições de patrimônio cultural, em que as ferramentas da forense digital e as técnicas estão cada vez mais entrando no campo da preservação. O projeto Vida Digital da Biblioteca Britânica[9] foi o primeiro a pesquisar o uso de ferramentas forenses para propiciar autenticidade em materiais digitais (John *et al*., 2010; John, 2008). Em 2010, o Conselho de Biblioteca e Recursos da Informação[10] publicou um relatório profícuo apresentando a aplicação da forense digital no setor do patrimônio cultural, estabelecendo firmemente um vínculo entre a curadoria digital e a forense digital (Kirschenbaum ***et al***, 2010). A forense digital e a arquivologia estão se juntando no desenvolvimento de ferramentas tais como *BitCurator*, um conjunto de ferramentas da forense digital na análise de dados para subsidiar a preservação digital em instituições de patrimônio cultural (wiki.bitcurator.net) e *Archivematica,* um conjunto de fluxos de trabalho de preservação digital segundo o ISSO-OAIS e as descobertas do InterPARES (www.archivematica.org). Na comunidade profissional da forense digital, a literatura tem sido predominantemente técnica,

[5] *Digital Records Forensics (DRF)* [N.T.].
[6] *School of Library, Archival and Information Studies (SLAIS)* [N.T.].
[7] *Faculty of Law* [N.T.].
[8] *Computer Forensics Division of Vancouver Police Department* [N.T.].
[9] *The Digital Lives at the British Library*. [N.T.].
[10] *Council on Library and Information Resources* [N.T.].

embora exista literatura propondo que a forense digital esteja situada dentro de um amplo arcabouço social e teórico (Mocas, 2001; Palmer, 2001, 2002; Pollitt, 2009). O campo da diplomática foi introduzido na literatura forense pelo cientista Fred Cohen, um pesquisador que neste momento faz parte do grupo InterPARES Trust da UBC (Cohen, 2011, 2012).

Este artigo está construído sobre trabalhos anteriores e apresenta uma exploração introdutória de um modelo de "documento de arquivo", o qual mapeia seus atributos tradicionais sobre uma análise técnica dos documentos de arquivo digitais. Este modelo está dividido em três camadas de abstração, a saber, a camada conceitual, a lógica e a física. Assim, propomos aplicar a diplomática de documentos digitais de arquivo que identifique os elementos essenciais de um documento de arquivo em cada camada de abstração. O projeto em andamento faz parte da minha pesquisa de doutorado e é um estudo dentro do projeto do InterPARES Trust (Documentos de Arquivo Confiáveis em uma Sociedade Cada Vez Mais Conectada)[12], Metadados***, Mutatis Mutandis:*** Parâmetros de ***Design*** para a Autenticidade na Nuvem e Entre Contextos (www.interpares.org)[13].

Uma visão técnica dos documentos de arquivo

A abstração é um processo de compreensão de objetos complexos ocultando todos os detalhes, exceto as caraterísticas essenciais de um conceito ou objeto específico necessário para concluir uma tarefa específica. Cada camada de abstração conterá seu próprio conjunto exclusivo de características e funcionalidades existentes, independentemente das outras camadas. Por exemplo, quando dirigimos um carro, precisamos conhecer a regras da estrada, como ligar o motor, como dirigir, como parar**,** etc. Não precisamos conhecer como funciona a combustão do motor ou como trabalha o sistema do computador de bordo. Essas são as camadas de abstração. Usando essa analogia em nossos sistemas digitais, precisamos entender como criar documentos de texto ou navegar pela internet, porém, não nos preocupamos com a programação que está por trás dessas tarefas.

Aos nossos olhos, um documento de arquivo existe como um objeto conceitual, como algo que pode ser impresso no papel a partir de nosso computador ou que pode ser legível na nossa tela. Sabemos que ele pode estar armazenado no dispositivo digital ou na nuvem, porém o usamos e pensamos na sua concepção visual. Nossa interação com ele é como um objeto conceitual – um documento de arquivo. Mas, precisamos entender os documentos de arquivo digitais enquanto objetos físicos e lógicos, para preservá-los e avaliar sua autenticidade.

Os documentos de arquivo digitais podem ser entendidos em três níveis de abstração **(Figura 1)**. No nível mais alto de abstração se encontra o objeto conceitual – o documento de arquivo como ele é reconhecido e entendido por uma pessoa. O documento de arquivo que olhamos no nosso monitor ou tela é apresentado para nós por um ***software, que é o*** aplicativo usado para criar o documento.

O documento de arquivo também existe como um objeto físico – uma inscrição de signos sobre um meio físico. A informação constituída por dados e as instruções para sua manipulação e apresentação são representadas em um de dois estados- ligado ou desligado – sinais eletrônicos- um sistema binário de 0s (desligado) e 1s (ligado). Um bit é a menor unidade de informação que pode ser armazenada ou manipulada; um byte (oito bits) é a unidade básica de armazenagem na memória de um computador.

Por último, o documento de arquivo existe como um objeto lógico- um objeto, ou provavelmente muitos objetos diferentes ou componentes digitais, reconhecidos e processados por um ***hardware*** e um *software* para a produção do documento de arquivo conceitual. É nesse nível que nossas analogias com o mundo analógico fracassam. Sabemos que um documento de arquivo pode ser representado por diferentes tipos de aplicativos de *software,* mas temos essencialmente a mesma apresentação (por exemplo, um documento em *Word* ou um documento em PDF) embora seja formado por componentes digitais diferentes, vinculados de formas diferentes pelos aplicativos e sistemas operacionais de *software* e *hardware* [2] (Trace, 2011; Thibodeau, 2002). O documento de arquivo pode também ser apresentado por diversos *softwares* e ter o mesmo conteúdo intelectual, porém, ter uma apresentação consideravelmente diferente (por exemplo, um

[12] *Trusting Records in an Increasingly Networked Society* [N.T.].
[13] *Metadata, Mutatis Mutandis: Design Requirements for Authenticity in the Cloud and Across Contexts* [N.T.].

documento em *Word* ou um documento em *.txt*).

A camada lógica inclui tudo o que acontece por trás da tela ("debaixo do capô").[14] Ela pode ser entendida como várias camadas de abstração **(figura 2)**. O nível mais alto inclui *softwares* aplicativos, como, por exemplo, um *software* de processamento de palavras – ou dados – programas de multimídia, aplicativos de bases de dados etc. Esses aplicativos são gerenciados pelo *software* do sistema – sistemas operacionais, *softwares* utilitários, compiladores etc.

Numa visão técnica de documentos de arquivo (digitais), isolamos coisas diferentes. Na diplomática tradicional, os elementos formais intrínsecos e extrínsecos existem tanto no conteúdo do documento como no meio físico. Na diplomática digital não podemos pensar da mesma forma. O contexto tecnológico do documento de arquivo inclui a linguagem em que os dados e as instruções do computador são representados e manipulados e os componentes físicos necessários para que um computador realize as tarefas predeterminadas. O meio físico não é mais um elemento extrínseco da forma, ele faz parte do contexto tecnológico no qual o documento de arquivo é criado, conservado e preservado (Duranti e Thibodeau, 2006).

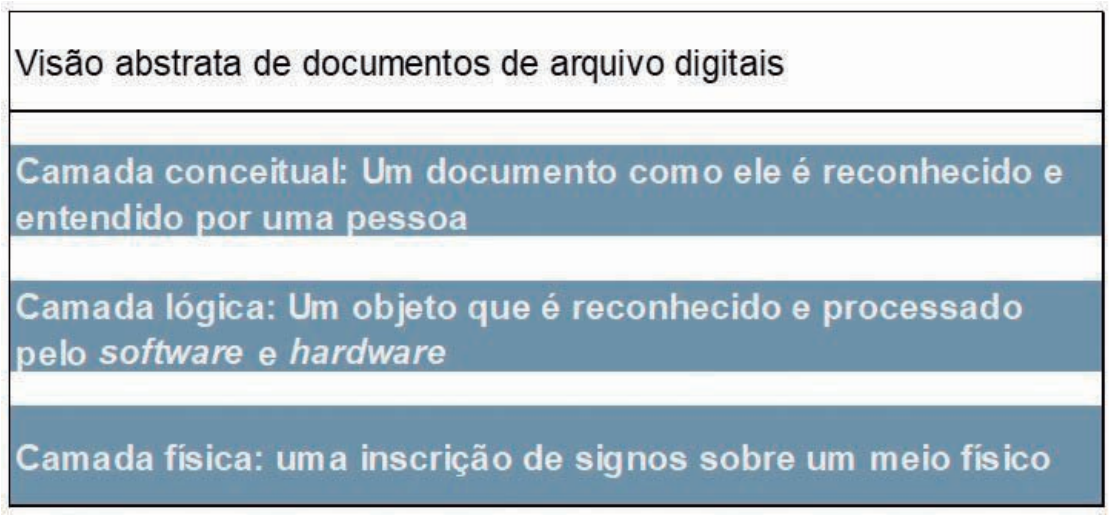

Figura 1. Uma visão abstrata de documentos de arquivo digitais

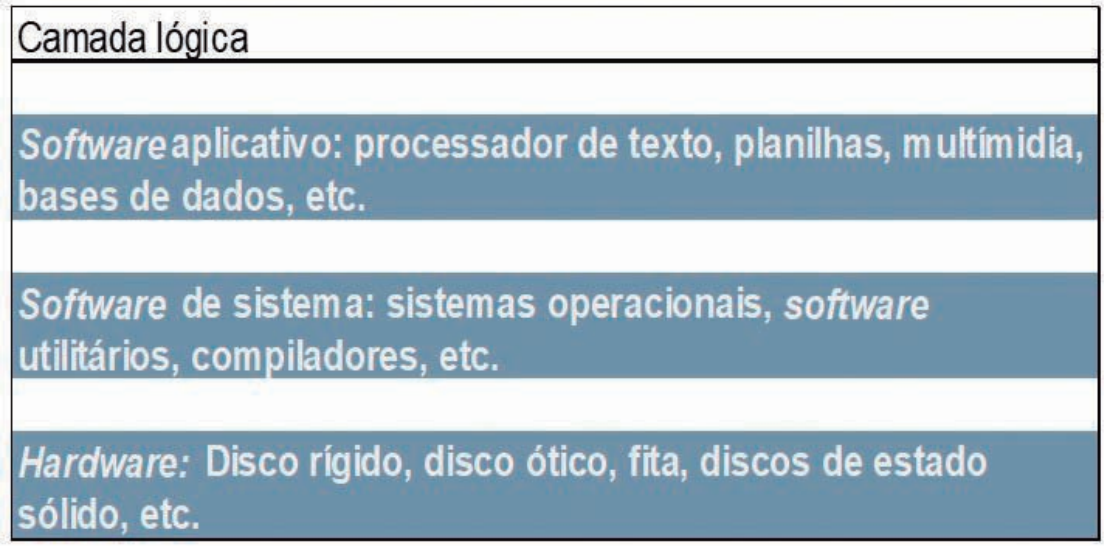

Figura 2. Camada lógica

[14] *("under the hood")*, no original [N.T.]

A maioria dos usuários interage com documentos de arquivo por meio dos ***softwares*** aplicativos. Eles são os aplicativos e sistemas operacionais que importam para a identificação de documentos de arquivo e outros materiais digitais que pretendemos analisar. A relação entre as camadas lógica e física é definida pelo aplicativo e o ambiente no qual esse aplicativo é executado no sistema operacional. Porém, a estrutura do nosso objeto conceitual é consideravelmente diferente da estrutura desse objeto conceitual nas camadas lógicas, e o mesmo conteúdo conceitual pode ser representado por diferentes aplicativos e, portanto, por diferentes codificações (Thibodeau, 2002). Onde encontramos os elementos formais extrínsecos e intrínsecos?

A forma documental

A forma documental é definida como as regras de representação de acordo com as quais o conteúdo de um documento de arquivo, seu contexto administrativo e documental e sua autoridade são comunicados através de seus elementos intrínsecos e extrínsecos. Os elementos formais extrínsecos determinam o layout e a aparência dos documentos de arquivo. Eles incluem elementos como a linguagem, os recursos de apresentação e símbolos especiais, e sua presença faz com que o documento possa ser executado. Os recursos extrínsecos também incluem anotações adicionadas durante a execução, manuseio ou o gerenciamento e a preservação do documento de arquivo. Os elementos intrínsecos transmitem a ação e seu contexto imediato nos quais o documento de arquivo se insere. Eles incluem nomes de pessoas, datas, localização, assunto, título, atestações e assim por diante, e sua presença faz com que o documento de arquivo fique completo (Duranti, 1998). Através da análise desses elementos, é possível estabelecer a identidade e rastrear a integridade do documento de arquivo ao longo do tempo, permitindo assim a avaliação da sua autenticidade.

No mundo do suporte em papel, esses elementos da forma são evidentes simplesmente com uma inspeção visual ao documento de arquivo. No ambiente digital, eles podem ser visíveis na manifestação conceitual do documento de arquivo, porém, pode existir também (ou não) na representação lógica do documento de arquivo. Assim, eles podem ser explícitos nos dados do documento de arquivo (conteúdo) ou capturados manualmente ou automaticamente como metadados associados ao documento de arquivo conceitual. Esses metadados podem estar imediatamente visíveis (por exemplo, o nome do arquivo e o caminho), visíveis por meio das funções disponíveis para qualquer usuário do sistema no qual o documento do arquivo é armazenado (por exemplo, o tamanho do arquivo a data de criação ou a última modificação), ou visível por meio simplesmente de uma investigação mais complexa utilizando ferramentas especializadas (forense digital).

Assim os metadados podem estar presentes nas camadas conceitual e lógica que formam o documento de arquivo. Ao identificar os seus elementos relacionados aos elementos formais extrínsecos e intrínsecos necessários para realizar uma análise diplomática, deve ser possível identificar documentos de arquivo em sistemas digitais e avaliar sua autenticidade. No entanto, existem vários desafios ao rastrear elementos diplomáticos nos metadados. Por exemplo: quem os cria? Quais são mantidos e quais são perdidos durante a transmissão, migração etc.? Quem é seu proprietário? Eles são acessíveis? Como eles estão vinculados com o documento de arquivo originário? Quais são necessários para serem preservados com o documento de arquivo ao longo do tempo? Essas são apenas algumas das perguntas que precisam ser feitas.

> Quando se trata de declarar documentos eletrônicos de arquivo, os profissionais de governança da informação vêm lutando há algum tempo com a questão de quais metadados preservar (Isaza, 2010).

Os arquivistas têm certa familiaridade com a classificação dos metadados por sua função ou propósito: descritivos, administrativos (técnicos, de direito, de preservação) e estruturais. Os metadados descritivos contêm elementos ou propriedades que identificam um documento de arquivo digital e subsidiam sua localização e interpretação. Os metadados administrativos são usados para gerenciar o documento de arquivo. Os metadados administrativos podem conter metadados que fornecem informação sobre o contexto técnico do documento de arquivo, informação sobre direitos e obrigações vinculados aos recursos digitais tais como a propriedade, o direito autoral ou outros direitos de propriedade intelectual, as restrições de uso e de segurança, e os metadados de preservação, descrevendo os requisitos para preservar o documento de arquivo ao longo do tempo e das mudanças tecnológicas. Os metadados estruturais documentam as relações estruturais entre ou dentro dos recursos digitais, como por exemplo, a estrutura de um arquivo dentro de um recurso digital ou a ligação entre páginas de internet em um ***Web Site***. Os metadados estruturais

auxiliam na exibição e no uso adequado de objetos complexos. As categorias dos metadados derivam da criação, conservação e preservação de recursos. Outra categoria pode ser identificada baseada no uso de recursos, tanto da analítica como do conteúdo gerado pelo usuário. O uso de metadados inclui elementos de dados ou propriedades coletados sobre ou dos usuários do documento de arquivo (por exemplo, ***tags*** sociais, registros de acesso, registro de busca de usuário). Os metadados também permitem a representação de material em vários níveis de agregações – da instituição arquivística, das séries ou coleções do documento de arquivo e da informação ou do componente digital (Zhang e Mauney, 2013).

Contudo, os metadados geralmente não são classificados de acordo com a função, mas de acordo com a fonte. Dada a prevalência de material digital apresentado como prova em litígios e tribunais e a falta de consistência com que os metadados são considerados pelos tribunais, a Conferência de Sedona[15] [3] oferece um exemplo relevante e importante: A Conferência de Sedona identifica sete tipos de metadados:

> (1) ***Metadados aplicativos***: São dados criados por um aplicativo específico para a informação armazenada eletronicamente (IEA)[16] sendo direcionados, embutidos no arquivo e movidos por este quando forem copiados; a cópia pode modificar os metadados aplicativos.
> (2) ***Metadados do documento***: Propriedades sobre o arquivo nele armazenado [...] como exemplo se incluem o autor do documento e a empresa, e as datas de criação e de revisão.
> (3) ***Metadados do e-mail:*** São dados armazenados no ***e-mail*** sobre o ***e-mail.*** Muitas vezes essa informação não é visível no aplicativo cliente do ***e-mail*** usado para criar o e-mail, como por exemplo, endereços de cópia oculta, data de recebimento [...].
> (4) ***Metadados embutidos***: Estão geralmente ocultos, mas são uma parte integrante da IEA, tais como "rastreio de modificações" ou "comentários" em um arquivo de processamento de texto [...]. Isso pode estar disponível somente no arquivo original ou inicial.
> (5) ***Metadados do sistema de arquivos***: são metadados geralmente gerados pelo sistema para rastrear a demográficos (nome, tamanho, lugar, tipo de uso, etc.) da IEA e que não esteja embutida, mas armazenada externamente a partir da IEA.
> (6) ***Metadados adicionados pelo usuário***: dados, possivelmente produto do trabalho, criados por um usuário quando copia, revisa e trabalha com um arquivo, incluindo anotações e informação subjetiva codificada.
> (7) ***Metadados adicionados pelo fornecedor***: são dados criados e mantidos pelo fornecedor eletrônico legalmente autorizado como resultado do procedimento do documento [...]; muito disso é usado para relatórios de processamento, cadeia de custódia e prestação de contas dos dados (The Sedona Conference, 2010).

Embora essa classificação possa ser útil com o objetivo de alertar os profissionais da área legal sobre as fontes de metadados associados a informações armazenadas eletronicamente, buscadas e compartilhadas na parte legal (***Discovery***), ela não descreve, nem prova a estrutura ou a forma documental do documento de arquivo, nem ajuda a avaliar sua autenticidade ou sua proveniência. De fato, de acordo com o guia de boas práticas para gerenciar documentos eletrônicos da ***Sedona***[17], "Na ausência de um requisito legal que seja contrário, as organizações não precisam preservar os metadados; mas pode ser útil preservá-los em alguns casos" (The Sedona Conference, 2007).

Nem sempre é papel dos metadados considerar a estrutura e a completude dos documentos de arquivo ou objeto digital que eles descrevem. Esse é o motivo dos esquemas de metadados orientados a objetivos que prescrevem um conjunto de elementos de metadados requeridos para um objetivo específico, tais como a recuperação da informação ou gerenciamento de coleções ou para a preservação. O que a diplomática digital tenta realizar é identificar todos os elementos necessários dos metadados que devem ser criados, gerenciados e preservados para identificar inequivocamente um documento de arquivo e demostrar sua integridade ao longo do ciclo de vida desse documento de arquivo, desde sua produção até sua gestão de uso, reuso e preservação. Note-se que isso não é um processo linear, é cíclico ou contínuo ao longo do tempo e acaba unicamente quando, ou se, o documento é destruído sem vestígios da sua existência.

[15] *The Sedona Conference* [N.T.].
[16] *Electronically Stored Information* ***(ESI***) [N.T.].
[17] *Sedona's best practice guidelines form managing electronic records* [N.T.].

Para identificar documentos de arquivo em sistemas digitais e avaliar sua autenticidade, devemos voltar aos princípios da diplomática arquivística. Devemos localizar os elementos formais extrínsecos e intrínsecos nos metadados de identidade e integridade, que podem ser localizados em várias camadas de abstração do documento de arquivo conceitual e lógico. O meio físico não é um problema. As regras que regem o objeto lógico são independentes de como os dados são escritos no meio físico. No nível lógico, a gramática (a interpretação dos ***bits***) é independente do meio físico. Uma vez os que os dados sejam lidos na memória, o tipo de meio e sua forma de inscrição não têm relevância. As regras que são aplicadas no nível lógico determinam como o fluxo de entrada é transformado dentro da memória do sistema e a sua saída como apresentação. O mapeamento do lógico para o físico pode ser alterado, sem afetar o nível logico **(Figura 3)**.

Podemos mapear os elementos formais intrínsecos da camada lógica dos documentos de arquivo? Os elementos dos metadados de identidade e de integridade, exigidos pelo InterPARES para avaliar a autenticidade dos documentos de arquivo digitais ao longo do ciclo de vida, foram identificados no Perfil do Aplicativo para Metadados de Autenticidade (IPAM)[18]. Esta pesquisa, seguindo o modelo de Singapura e a Iniciativa de Metadados da Dublin Core (DCMI), desenvolveu um conjunto de requisitos funcionais, os quais foram modelados através de diagramas de entidades-relações. Estes requisitos funcionais determinaram um desenvolvimento de metadados, a saber:

> [...] necessários e suficientes para apoiar a presunção de autenticidade dos documentos de arquivo, interoperáveis entre sistemas e no decorrer do tempo, adequados para a descrição de arquivos e úteis tanto para a recuperação como para exibição significativa de documentos de arquivo.

Elementos extrínsecos da forma - camada lógica

	Aplicação (*Software*)	Aplicação (Usuário)	Sistema	*Hardware*
Meio	Formato do arquivo		Localização	Disco rígido, disco ótico etc.
Escrita	Formato do arquivo		NA	bits
Linguagem	Codificação de caracteres		NA	NA
Símbolos especiais e autenticações	Podem ser arrastados para o arquivo	Assinaturas digitais	*Login* , Informação da conta do usuário	NA
Anotações	Incorporadas ou vinculadas		NA	NA

Figura 3

Como foi comentado anteriormente pelo estudo do InterPARES, os elementos dos metadados foram codificados por função, funções essas identificadas de acordo com os princípios da arquivologia e da teoria da diplomática. O perfil do aplicativo captura, assim, a prova dos anexos; autenticação; ligações com outros documentos de arquivo com os quais o documento de arquivo se relaciona e seu contexto, datas e os momentos em que é necessário documentar o ciclo de vida do documento de arquivo no ciclo de vida; ligações com a documentação externa que rege a preservação, transferência e acesso ao(s) documento(s) ao longo do tempo; regras de apresentação que determinam a aparência de uma entidade; manipulação da informação; informação da localização do armazenamento, cópias de segurança (***backups)*** ou duplicação; identificação de indivíduos ou entidades legalmente definidas, que sejam sujeitos de direitos e deveres e que sejam reconhecidos pelo sistema jurídico, com a capacidade ou o potencial de atuar legalmente em relação com o(s) documento(s) de arquivo, as restrições ou privilégios de direitos e os acessos que se aplica(m) ao(s) documento(s) de arquivo, o assunto da ação ou matéria que pertence ao(s) documento(s) de arquivo e a informação tecnológica sobre o(s) transmissor(es) da forma e o conteúdo do documento de arquivo ao longo do tempo (Tennis e Rogers, 2012ª, 2012b).

[18] *Application Profile for Authenticity Metadata (IPAM)* [N.T.].

Próximos passos

Este modelo de metadados conceitual baseado nos princípios da diplomática, pode agora ser mapeado dentro das camadas lógicas do documento de arquivo digital. Para isso, será necessário analisar os metadados existentes que são ou podem ser inseridos pelos ***agentes*** que manipulam o documento de arquivo, isto é, as pessoas e os sistemas que inserem metadados manual ou automaticamente. Na interface entre a camada conceitual e a camada do aplicativo (camada lógica), estão as pessoas que participam da sua criação e recriação ao longo do tempo, as pessoas que participam em seu gerenciamento e/ou sua preservação e as pessoas e os sistemas que usam o documento de arquivo. Na camada lógica do aplicativo e do ***software*** e ***hardware*** do sistema, estão os elementos dos sistemas que o gerenciam.

Assim, vamos voltar para a forense digital como foi mencionado antes. Peritos da forense digital estão localizando e usando metadados (que é fundamental para análise de dados e visualização, segurança, inteligência e vigilância) para reconstruir eventos criminais ou incidentes de segurança. Consideremos, por exemplo, os metadados associados com um ***tweet*** **(tabela 1)**:

Campo de metadados	Descrição	Elemento diplomático
Created_at	Registro de data e hora UTC para criação de *tweets*	Data (Cronológica)
User_ID	Identificação da postagem de um *tweet*	Pessoas (autor)
Handle	Nome de tela do usuário (diferente do nome de usuário)	Pessoas (autor)
Retweet_ID	Identificação de quem postou um *retweet*	Vínculo Arquivístico
Retweet_user	Nome de usuário da pessoa que *retweetou*	Pessoas (autor)
Geo_enabled	Geolocalização disponível do usuário (opcional)	Data (tópica)
Place	Geolocalização de onde o usuário *tweetou*	Data (tópica)
Coordinates	Coordenada de geolocalização desde onde o *tweet* foi enviado	Data (tópica)
In_reply_to_user_ID	Identificação única do usuário que respondeu	Pessoas (destinatário)
Source	Aplicativo usado no *tweet* ou mensagem direta (por exemplo, de um iPhone ou de um aplicativo específico de *Twitter)*	Contexto tecnológico
Follow_request_sent	Indica solicitação para seguir o usuário	Anotação

Fonte: Patzakis, 2012 Tabela I.

Tabela 1

O campo dos metadados identificados na tabela 1 e relatados por meio da investigação da forense digital, são guardados automaticamente por ***Twitter*** e podem ser interpretados como conceitos diplomáticos. Porém, os metadados não são à prova de falhas dos eventos.

Segue uma captura de tela das propriedades de um documento de ***Word*** indicando que sou a autora. Ela é gerada automaticamente como resultado da configuração do meu notebook (um iMac executando a versão OSX 10.9.2) **(figura 4)**.

Figura 4. Propriedade do Documento – Word (1)

É possível adicionar mais metadados no documento de arquivo em questão. Na **Figura 5** foram inseridos título, assunto e palavras chaves.

Esses elementos de metadados, porém, não são necessariamente elementos formais intrínsecos confiáveis já que podem ser facilmente modificados **(figura 6)**.

Além disso, o mesmo documento de arquivo salvo em outro formato não mostra os mesmos metadados **(figura 7)**.

A pesquisa continua em andamento atualmente pelo InterPARES Trust, no ***design*** de sistemas para metadados em todas as camadas de abstração para que possam ajudar na presunção de autenticidade de documentos de arquivo digitais.

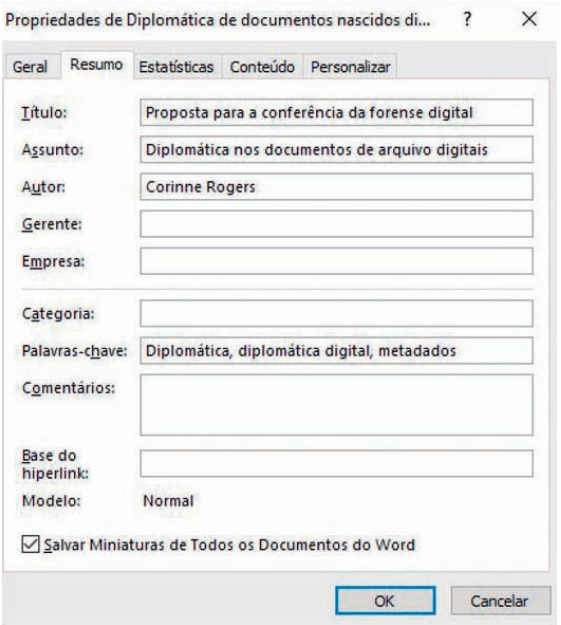

Figura 5. “Propriedades do Documento- Word (2)”

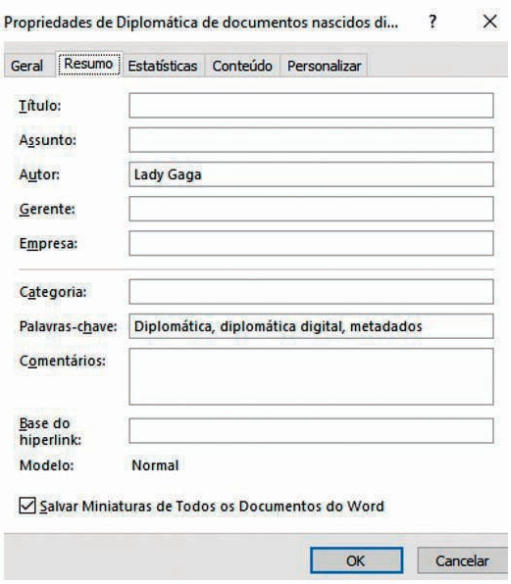

Figura 6. Propriedade do Documento (3)

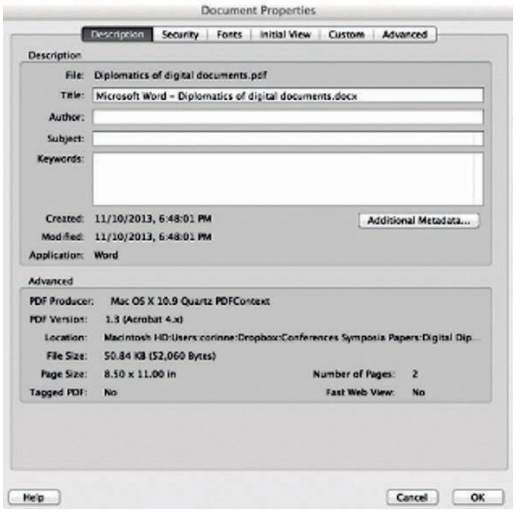

Figura 7. Propriedade do Documento- PDF

Conclusões

Os documentos de arquivo digitais estão estruturados em: dados (conteúdo) gerados pelo usuário, metadados, gerados pelo sistema, que identificam a sua fonte e a sua localização, metadados gerados por aplicativos que gerenciam a aparência e o desempenho do documento de arquivo (como, por exemplo, o formato do arquivo original), metadados gerados por aplicativos que descrevem os dados (como por exemplo os metadados do sistema operacional em que o arquivo foi gerado), e metadados que descrevem os dados, gerados pelo usuário. A diplomática digital, baseada nos princípios da diplomática tradicional, pode subsidiar na identificação de documentos de arquivo digitais por meio de seus metadados e determinar quais metadados são necessários para serem capturados, gerenciados e preservados.

Notas do texto original

Os primeiros profissionais se referiram à prática de computação forense. Como os dispositivos de armazenamento digital se tornaram onipresentes e não eram necessariamente computadores tradicionais, o

termo "digital" começou a substituir o de "computador" (Whitcomb, 2002).

Para informação mais detalhada sobre este modelo tripartido de documento, ver Thibodeau (2002), Trace (2011) "Overview of Technological Approaches to DigitalPreservation and Challengesin Coming Years 1", in ***The State of Digital Preservation: Na International Perspective***, CLIR, www.clir.org/pubs/reports/pub107/thibodeau.html. Para maior informação sobre computadores e documentos de arquivo digitais, ver Ciaran Trace (2011) "Beyond the Magic to the Mechanism: Computers, Materiality, and What It Means for Records to Be 'Born Digital'", ***Archivaria***, Vol. ***72***, December, pp. 5-27.

A Conferência de Sedona (https://thesedonaconference.org) é um instituto de pesquisa e de educação não partidário, dedicado ao avanço da lei e das políticas nas áreas de direito antitruste, litígios complexos e direitos de propriedade intelectual. Ela produziu vários relatórios influentes sobre ***e-discovery*** e prova eletrônica. O trabalho da Conferência de Sedona foi influente em caso de lei decidindo sobre prova digital. Veja, por exemplo, ***Aguilar v. Immigration & Customs Enforcement Div. of US Dep't of Homeland Sec.***, 255 F.R.D. 350 (S.D.N.Y. 2008) dedicado à classificação e à importância dos metadados, citado por (the Sedona Conference (2013)).

Referências:

Boyle, L. (1992), "Diplomatics", in Powell, J.M. (Ed.), ***Medieval Studies: An Introduction***, SyracuseUniversity Press, Syracuse, New York, NY, pp. 82-113.

Cohen, F. (2011), ***Digital Forensic Evidence Examination***, 3rd ed., Fred Cohen and Associates,Livermore, CA.

Cohen, F. (2012), ***The Future of Digital Forensics***, presented at the Trust and Conflicting Rights inthe Digital Environment, Vancouver, BC.

Diamond, E. (1994), "The archivist as forensic scientist – seeing ourselves in a different way", ***Archivaria***, Vol. 38 (Fall), pp. 139-154.

Duranti, L. (1998), ***Diplomatics: New Uses for an Old Science***, Scarecrow Press, Lanham, MD.

Duranti, L. (2005), ***The Long-term Preservation of Authentic Electronic Records: Findings of theInterPARES Project***, Archilab, San Miniato.

Duranti, L. (2009), "From digital diplomatics to digital records forensics", ***Archivaria***, Vol. 68(Fall), pp. 39-66.

Duranti, L. (2010), ***Digital records Forensics: Continuing in Mabillon's Footsteps***, presented at theAssociation of Canadian Archivists, Halifax, NS.

Duranti, L. and Endicott-Popovsky, B. (2010), "Digital records forensics: a new science andacademic program for forensic readiness", ***Journal of Digital Forensics, Security and Law***,Vol. 5 No. 2, pp. 1-12, available at: www.jdfsl.org/subscriptions/JDFSL-V5N2-Duranti.pdf

Duranti, L. and MacNeil, H. (1997), "The preservation of the integrity of electronic records: anoverview of the UBC-MAS research project", ***Archivaria***, Vol. 42 (Spring), pp. 46-67.

Duranti, L. and Preston, R. (2008), ***Research on Permanent Authentic Records in ElectronicSystems (InterPARES) 2: Experiential***, Interactive and Dynamic Records, AssociazioneNazionale Archivistica Italiana, Padova.

Duranti, L. and Thibodeau, K. (2006), "The concept of record in interactive, experiential anddynamic environments: the View of InterPARES", ***Archival Science***, Vol. 6 No. 1, pp. 13-68.

Irons, A. (2006), "Computer forensics and records management – compatible disciplines", ***Records Management Journal***, Vol. 16 No. 2, pp. 102-112.

Isaza, J. (2010), ***Metadata in Court: What RIM, Legal and IT Need to Know***, ARMA International Education Foundation, Pittsburgh, PA, p. 26.

John, J.L. (2008), "Adapting existing technologies for digitally archiving personal lives: digital forensics, ancestral computing, and evolutionary perspectives and tools", ***Proceedings of The Fifth International Conference on Preservation of Digital Objects***, ***The British Library, London***, pp. 46-55, available at: www.bl.uk/ipres2008/ipres2008-proceedings.pdf

John, J.L., Rowlands, I., Williams, P. and Dean, K. (2010), "Digital lives: personal digital archives for the 21st century – an initial synthesis (Digital Lives Research Paper)", available at: http://britishlibrary.typepad.co.uk/files/digital-lives-synthesis02-1.pdf

Kirschenbaum, M.G., Ovenden, R. and Redwine, G. (2010), ***Digital Forensics in Born Digital Cultural Heritage Collections***, Council on Library and Information Resources, Washington, DC.

MacNeil, H. and Gilliland-Swetland, A. (2005), "Authenticity task force report", in Duranti, L. (Ed.), ***The Long-term Preservation of Authentic Electronic Records: Finding of the InterPARES Project***, Archilab, San Miniato, Italy.

Mocas, S. (2004), "Building theoretical underpinnings for digital forensics research", ***Digital Investigation***, Vol. 1 No. 1, pp. 61-68.

Palmer, G. (2001), "A road map for digital forensic research (DFRWS technical report)", availableat: www.dfrws.org/2001/dfrws-rm-final.pdf

Palmer, G. (2002), "Forensic analysis in the digital world", ***International Journal of DigitalEvidence***, Vol. 1 No. 1, available at: www.ijdc.org/docs/IJDE_1.3.doc

Patzakis, J. (2012), "Key Twitter and Facebook metadata fields forensic investigators need to beaware of", ***Forensic Focus – Articles***, available at:http://articles.forensicfocus.com/2012/04/25/key-twitter-and-facebook-metadata-fields-forensic-investigators-need-to-be-awareof/ (accessed 1 August 2012).

Pollitt, M.M. (2009), "Digital Forensics as a Surreal Narrative", in Peterson, G. and Shenoi, S. (Eds),***Advances in Digital Forensics***, Springer Berlin Heidelberg, Heidelberg, Vol. 306, pp. 3-15,available at: http://link.springer.com/10.1007/978-3-642-04155-6_1

Tennis, J.T. and Rogers, C. (2012a), "Authenticity metadata and the ipam: progress toward theInterPARES application profile", ***Proceedings of the International Conference on DublinCore and Metadata Applications***, ***DCMI, Kuching, Sarawak***, pp. 38-45, available at: http://dcevents.dublincore.org/index.php/IntConf/dc-2012/schedConf/presentations

Tennis, J.T. and Rogers, C. (2012b), ***General Study 15: Metadata Application Profiles forAuthenticity***, University of British Columbia, Vancouver, BC.

The Sedona Conference (2007), "The Sedona principles: best practice guidelines & commentary formanaging information & records in the electronic age", 2nd ed., The Sedona Conference,available at:http://find.galegroup.com/gtx/infomark.do?&contentSet_IAC-Documents&type_retrieve&tabID_T002&prodId_LT&docId_A130739169&source_gale&srcprod_LT&userGroupName_ubcolumbia&version_1.0

The Sedona Conference (2010), ***The Seconda Conference Glossary: E-Discovery & DigitalInformation Management***, 3rd ed., The Sedona Conference, Sedona, AZ.

The Sedona Conference (2013), ***The Sedona Conference Commentary on Ethics & Metadata***, TheSedona Conference, Sedona, AZ.

Thibodeau, K. (2002), "Overview of technological approaches to digital preservation andchallenges in coming years 1", ***The State of Digital Preservation: An InternationalPerspective***, CLIR, available at: www.clir.org/pubs/reports/pub107/thibodeau.html

Trace, C. (2011), "Beyond the magic to the mechanism: computers, materiality, and what it meansfor records to be 'Born Digital'", ***Archivaria***, Vol. 72, pp. 5-27, available at: http://journals.sfu.ca/archivar/index.php/archivaria/article/view/13358/14660

Whitcomb, C.M. (2002), "An historical perspective of digital evidence: a forensic scientist's view",***International Journal of Digital Evidence***, Vol. 1 No. 1, available at: www.utica.edu/academic/institutes/ecii/publications/articles/9C4E695B-0B78-1059-3432402909E27BB4.pdf

Zhang, J. and Mauney, D. (2013), "When archival description meets digital object metadata: atypological study of digital archival representation", ***The American Archivist***, Vol. 76 No. 1, pp. 174-191.