

# O DESAFIO DA MANUTENÇÃO DA INTEGRIDADE DE DOCUMENTOS ARQUIVÍSTICOS DIGITAIS

Cintia Aparecida Chagas (Universidade Federal de Minas Gerais)  
Fábio Lopes de Andrade (Universidade Federal de Minas Gerais)

## 1 INTRODUÇÃO

Os documentos arquivísticos digitais diferem dos documentos arquivísticos em papel significativamente. Eles são voláteis e sujeitos à perda, à alteração intencional ou não, à contaminação ou corrupção, mesmo quando ainda estão sob custódia de seu criador (ROGERS, 2016).

A integridade de um arquivo surge da garantia de que ele não foi alterado em qualquer forma não autorizada ou indocumentada; todo o propósito da preservação digital é manter o arquivo como ele foi originalmente criado e mantido. Possíveis ameaças à integridade de um objeto incluem corrupção acidental, alteração deliberada por um usuário não autorizado e alteração causada por código malicioso, tal como um vírus. A fim de proteger a integridade de um objeto digital, deve ser possível confirmar que ele não foi alterado de qualquer forma não autorizada como, por exemplo, através de corrupção, eliminação ou adição (MILLAR, 2009).

Neste sentido, o presente trabalho investigou como fenômenos físicos, relacionados ao decaimento magnético espontâneo dos suportes carreadores magnetizados, afetam a camada magnética de um disco rígido típico. Por meio

de uma revisão bibliográfica, foram identificados os fatores envolvidos na deterioração dos *bits*, de forma a comprometer a integridade da informação arquivística digital. Ao final, buscou-se demonstrar como os repositórios arquivísticos digitais confiáveis (RDC-Arqs) verificam a integridade dos pacotes de informação, por meio de algoritmos criptográficos, no processo de *Ingest*.

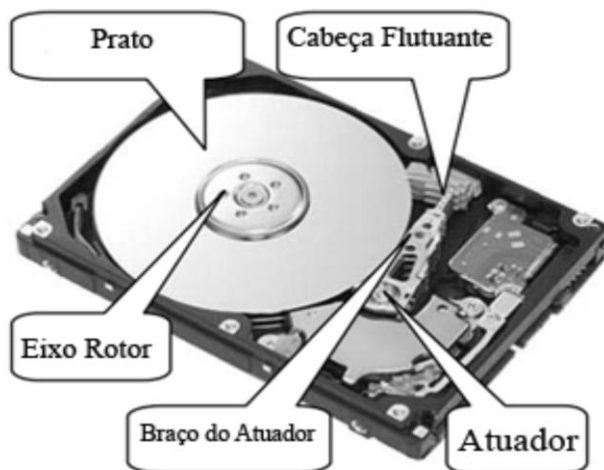
## 2 O FUNCIONAMENTO DE UM *DRIVE* DE DISCO RÍGIDO TÍPICO

As unidades de disco se tornaram, segundo Yuhui Deng (2011), o mais importante meio de armazenamento persistente que oferece alto desempenho, grande capacidade e alta confiabilidade. Eles são os principais dispositivos de armazenamento utilizados em ambientes computacionais altamente dinâmicos e em constante mudança. Em um *drive* de disco rígido, os dados binários são armazenados em uma fina camada magnética, chamada de meio de gravação, depositada em um substrato (JIAN-GANG ZHU, 2003).

Um disco magnético é composto de um ou mais pratos de alumínio, vidro ou cerâmica com um revestimento magnetizável (TANENBAUN, 2006).

A Figura 1 mostra a arquitetura típica de uma unidade de disco rígido. Ela consiste principalmente de pratos, eixo rotor, braço do atuador, cabeça flutuante, atuador.

Figura 1 - Arquitetura de um Disco Rígido



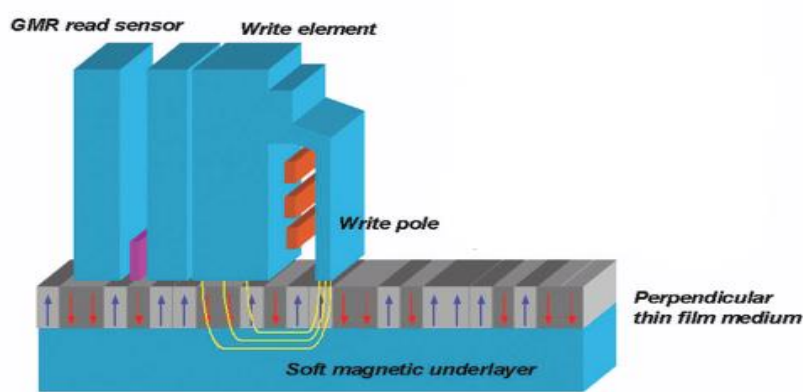
Fonte: Yuhui, 2011. Adaptado. Tradução nossa.

Uma cabeça (ou cabeçote) flutuante contém uma bobina de indução que flutua logo acima da superfície, apoiado sobre um colchão de ar. Quando

uma corrente elétrica passa pelo cabeçote, ele magnetiza a superfície logo abaixo dele, alinhando as partículas magnéticas para a esquerda ou para a direita, dependendo da polaridade da corrente (TANENBAUM, 2006).

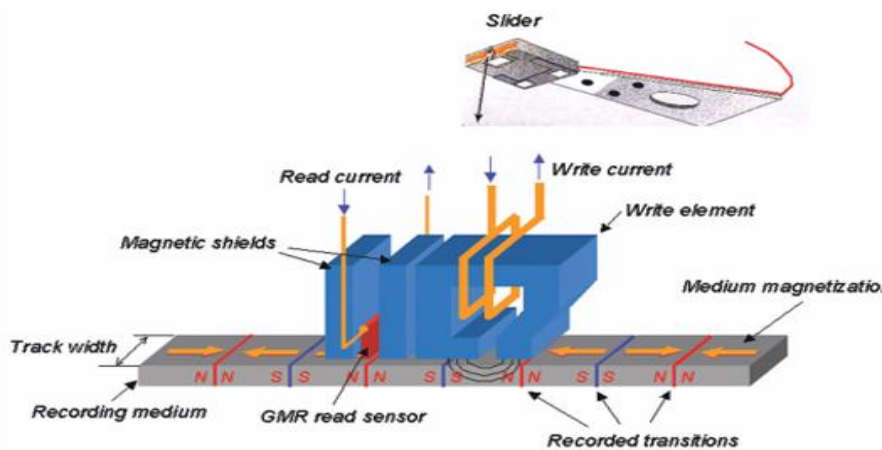
Quando o cabeçote passa sobre uma área magnetizada, uma corrente positiva ou negativa é induzida nele, o que possibilita a leitura dos *bits* armazenados antes. Assim, à medida que o prato gira sob o cabeçote, um conjunto de *bits* pode ser escrito e, mais tarde, lido (TANENBAUM, 2006).

**Figura 2** - Esquema de gravação perpendicular. A camada inferior macia atua como um eficiente caminho de fluxo de campo de escrita e, efetivamente, se torna parte da cabeça de escrita.



Fonte: JIAN-GANG ZHU, 2003, Adaptado.

**Figura 3** - Esquema de gravação longitudinal utilizado nos discos rígidos atuais<sup>106</sup>



Fonte: JIAN-GANG ZHU, 2003, Adaptado.

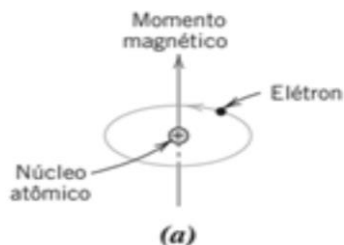
<sup>106</sup> . A cabeça de gravação consiste em um elemento de escrita indutivo e em um elemento de leitura magneto resistivo gigante (GMR).

### 3 O MAGNETISMO ENVOLVIDO NA GRAVAÇÃO DOS *BITS*

As propriedades magnéticas macroscópicas dos materiais são consequência dos momentos magnéticos que estão associados aos elétrons individuais. Cada elétron, em um átomo, possui momentos magnéticos que se originam de duas fontes, de acordo com Callister e Rethwisch (2016):

Uma fonte está relacionada com seu movimento orbital ao redor do núcleo; sendo uma carga em movimento, um elétron pode ser considerado um pequeno circuito circular com corrente, que gera um campo magnético muito pequeno e que apresenta um momento magnético ao longo do seu eixo de rotação, como está ilustrado esquematicamente na Figura 4.

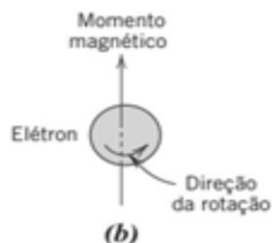
**Figura 4** - Momento magnético associado a um elétron em órbita



Fonte: CALLISTER e RETHWISCH, 2016. P.744. Adaptado.

Cada elétron também pode ser considerado como se estivesse girando ao redor de um eixo; o outro momento magnético tem sua origem nessa rotação do elétron, e está direcionado ao longo do eixo de rotação, como mostra a Figura 5. Os momentos magnéticos de *spin* podem estar apenas em uma direção “para cima” ou em uma direção antiparalela, “para baixo”. Dessa forma, cada elétron em um átomo pode ser considerado como se fosse um pequeno ímã, que possui momentos magnéticos permanentes, orbital e de rotação (CALLISTER e RETHWISCH, 2016).

**Figura 5** - Demonstração do momento magnético associado a um elétron girando em torno de seu eixo

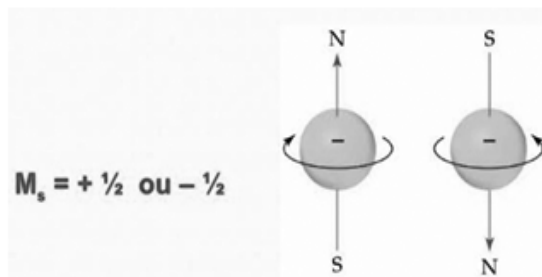


Fonte: CALLISTER e RETHWISCH, 2016. P.744. Adaptado.

Em cada átomo individual, os momentos orbitais de alguns pares eletrônicos se cancelam mutuamente; isso também é válido para os momentos de *spin*. Por exemplo, o momento de *spin* de um elétron com *spin* para cima cancelará aquele de um elétron com *spin* para baixo. O momento magnético resultante de um átomo é, então, simplesmente a soma dos momentos magnéticos de cada um dos seus elétrons constituintes, incluindo tanto as contribuições orbitais quanto as de *spin*, e levando em consideração os cancelamentos de momentos (CALLISTER e RETHWISCH, 2016).

O número quântico rotatório, também chamado *spin*, diz respeito ao movimento de rotação do elétron em torno do seu eixo. O movimento do elétron ao redor do núcleo atômico gera um campo magnético externo. Por outro lado, o movimento de rotação do elétron em torno do seu eixo gera outro campo magnético. A mecânica quântica estabelece que a interação desses dois campos magnéticos é quantizada e são possíveis apenas dois estados. Esses dois campos magnéticos ou se orientam paralelamente e no mesmo sentido ou paralelamente e em sentidos opostos, conforme exemplificado na Figura 6. Foram introduzidos os números quânticos  $+ \frac{1}{2}$  e  $- \frac{1}{2}$  para os dois spins possíveis, denominados *spin* paralelo e *spin* antiparalelo (CALLISTER e RETHWISCH, 2016).

Figura 6 - Representação de números quânticos rotativos



Fonte: Elaborado pelos autores.

#### 4 O “TRILEMA” DA GRAVAÇÃO MAGNÉTICA E A PERDA DE INTEGRIDADE DAS CADEIAS DE BITS

O registro magnético se baseia no armazenamento de informações em mídias compostas de grãos de materiais com uma alta anisotropia<sup>107</sup> magneto cristalina. Os grãos podem ser considerados como sistemas biestáveis capazes de representar *bits* de informação em termos da polaridade dos grãos. A

<sup>107</sup> As propriedades físicas dos monocristais de algumas substâncias dependem da direção cristalográfica na qual as medições são feitas (CALLISTER e RETHWISCH, 2016).

estabilidade das informações é fornecida por uma barreira energética de anisotropia KV, onde K é a constante de anisotropia e V corresponde ao volume de grãos. Há muito se percebeu que o fenômeno do superparamagnetismo<sup>108</sup> (SPM) define o limite superior da densidade de armazenamento de informações no registro magnético. O SPM surge devido à ativação térmica que pode causar a comutação espontânea da magnetização sobre a barreira energética da anisotropia, com a consequente perda de informações gravadas (EVANS et al., 2011).

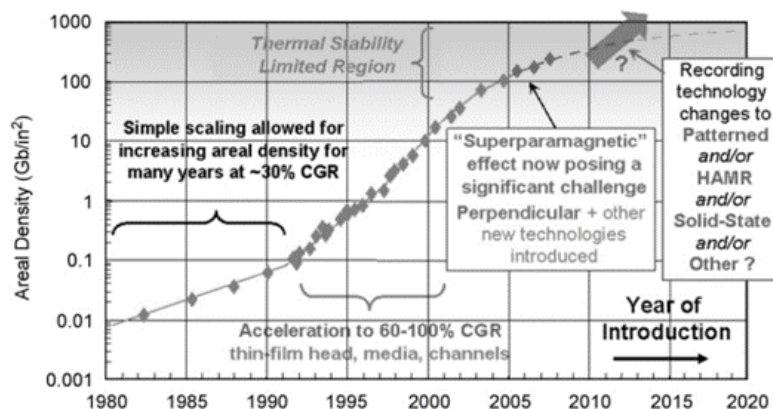
O armazenamento de dados magnéticos é difundido na preservação de informações digitais e o ritmo acelerado do desenvolvimento de computadores exige cada vez mais capacidade, conforme pode ser visualizado na Figura 7. Aumentar a densidade de armazenamento para discos rígidos magnéticos requer um tamanho de *bit* reduzido, antes considerado limitado pela estabilidade térmica dos grãos magnéticos constituintes. A densidade de armazenamento limitada na gravação magnética é investigada tratando a escrita de *bits* como um processo termodinâmico. É introduzido um fator de "gravabilidade térmica" e é mostrado que as densidades de armazenamento serão limitadas a 15 a 20 TBit/in<sup>2</sup> a menos que a tecnologia possa ir além das magnitudes do campo de gravação atualmente disponíveis (EVANS et al., 2011).

A fim de atender à crescente demanda por densidade de armazenamento e miniaturização de dispositivos de gravação magnética, o tamanho da célula de armazenamento magnético continua a diminuir, o que leva a uma inversão mais fácil dos domínios magnéticos. A quantificação do efeito secundário térmico magnético, que eventualmente levará à inversão da magnetização, é de importância crucial para prever a confiabilidade temporal das informações armazenadas em dispositivos de gravação magnética e para manter tolerâncias altamente precisas em sensores magnéticos (GU et al., 2011).

---

<sup>108</sup> Fenômeno que acompanha a diminuição no tamanho de uma nanopartícula — as suas propriedades físicas e químicas sofrem mudanças drásticas; além disso, o grau de mudança depende do tamanho da partícula (ou seja, do número de átomos) (CALLISTER e RETHWISCH, 2016).

Figura 7 - Crescimento das densidades de área para gravação convencional<sup>109</sup>

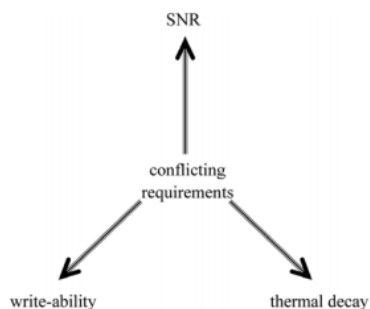


Fonte: Wood, 2008.

Mais apropriadamente, a deterioração térmica e a capacidade de escrita devem ser consideradas como restrições ao espaço de projeto. Em qualquer caso, deixa-se uma escolha entre "opções igualmente indesejáveis", que é, por definição, um dilema e, de certa forma, brincadeira que os funcionários da Empresa Seagate, produtora de drives de disco rígido mundialmente conhecida, chamaram de "trilema" da gravação magnética, exemplificado na Figura 8, para expressar que há três escolhas entre as opções indesejáveis. Neste sentido, qualquer avanço na evolução da gravação magnética pode ser considerado como um meio de quebrar o trilema ou, mais precisamente, adiar o trilema (RICHTER, 2007).

<sup>109</sup> Densidade de área de drives de disco rígido vs. ano de introdução. A taxa de aumento na densidade da área está diminuindo significativamente. Para a tecnologia de gravação convencional, as questões fundamentais forçam as contrapartidas entre: 'gravabilidade', 'relação sinal-ruído (SNR - *signal to-noise ratio*)' e 'estabilidade térmica'. Acredita-se que o limite da gravação convencional seja em torno de 1 Terabit / polegada quadrada.

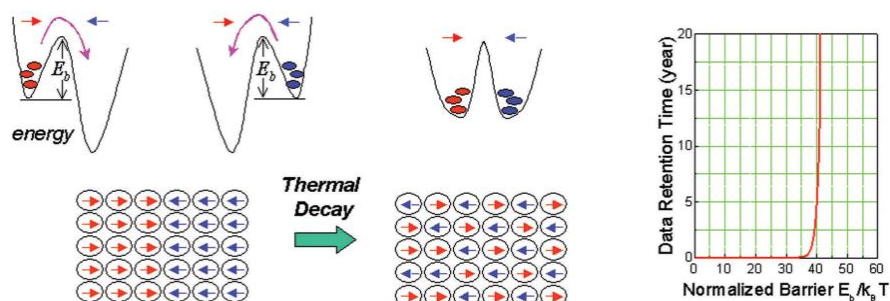
Figura 8 - O trilema da gravação magnética<sup>110</sup>



Fonte: RICHTER, 2007.

A Figura 9 ilustra uma transição em estado de ausência de equilíbrio térmico (esquerda) e a assimetria entre os dois poços de energia, que é causada pelo campo de desmagnetização gerado na transição. A excitação térmica produzirá uma reversão espontânea da magnetização dos grãos: quando o equilíbrio térmico for alcançado, a transição é completamente apagada. Em outras palavras, o tempo de retenção de dados em um meio de gravação é sempre finito e é apenas uma questão de tempo até que os dados sejam termicamente apagados.

Figura 9 - Apagamento espontâneo de uma transição registrada sob excitação térmica.



Fonte: JIAN-GANG Zhu, 2003.

## 5 VERIFICAÇÃO DE INTEGRIDADE DE OBJETOS DIGITAIS POR MEIO DE ALGORITMOS COMPUTACIONAIS

<sup>110</sup> Attingir alta relação sinal-ruído (*signal to-noise ratio* - SNR) requer a utilização de grãos pequenos. Grãos pequenos têm barreiras energéticas (KV) que são muito pequenas para garantir a estabilidade das informações registradas. Um aumento na densidade anisotrópica de energia (K) aumenta o campo de escrita necessário para além da capacidade dos materiais dos cabeçotes disponíveis

A fixidez, de acordo com NSDA (2014), é propriedade de um arquivo permanecer fixo ou inalterado. Neste sentido, é sinônimo de integridade em nível de *bits*.

As funções de *hash* criptográfico<sup>111</sup> se tornaram muito populares e hoje são utilizadas em muitos domínios diferentes que requerem troca de informações. Na maioria das vezes, são aplicadas na segurança informação digital, em operações como transações bancárias, verificação de assinaturas digitais, autenticação em geral e o envio de quaisquer dados via Internet, com o objetivo de proporcionar um nível muito elevado segurança da troca de informações (PAMULA e ZIEBINSKI, 2009).

Um dos métodos mais simples para verificação de fixidez é usar os *checksums*<sup>112</sup>. Um *checksum* é um algoritmo executado em um arquivo<sup>113</sup> cujo número resultante é referido como o valor do *checksum*. O valor do *checksum* é um número único atribuído a um arquivo com base no conteúdo do arquivo no nível do *bit*. Depois que este valor de *checksum* é conhecido, ele pode ser recalculado no arquivo novamente em qualquer tempo. Comparando os valores do checksum, se os números forem idênticos, então o arquivo não mudou; se o mesmo número não foi gerado, então significa que o arquivo mudou de alguma forma. Se diferenças são detectadas, este método não lhe diz o que mudou ou quando foi que algo mudou, apenas mostra que os dois arquivos não são mais idênticos no nível do bit (KUSSMANN, 2012).

Os *checksums* têm três usos principais:

- Saber que um arquivo foi recebido corretamente de um proprietário de conteúdo ou fonte e depois transferido com sucesso para o armazenamento de preservação.
- Saber que a fixidez do arquivo foi mantida quando esse arquivo está sendo armazenado.
- Ser compartilhado com os usuários do arquivo no futuro para que eles saibam que o arquivo foi corretamente recuperado do armazenamento e entregue a eles (DIGITAL PRESERVATION COALIZATION, 2020).

Um algoritmo é, de acordo com Cormen et al. (2012), qualquer procedimento computacional bem definido que toma algum valor ou conjunto de valores como entrada e produz algum valor ou conjunto de valores como saída. Portanto, um algoritmo é uma sequência de etapas computacionais que transformam a entrada na saída.

---

<sup>111</sup> Cryptographic hash functions. Tradução nossa.

<sup>112</sup> Soma de verificação. Tradução nossa.

<sup>113</sup> O termo arquivo, neste caso, é sinônimo de *file*.

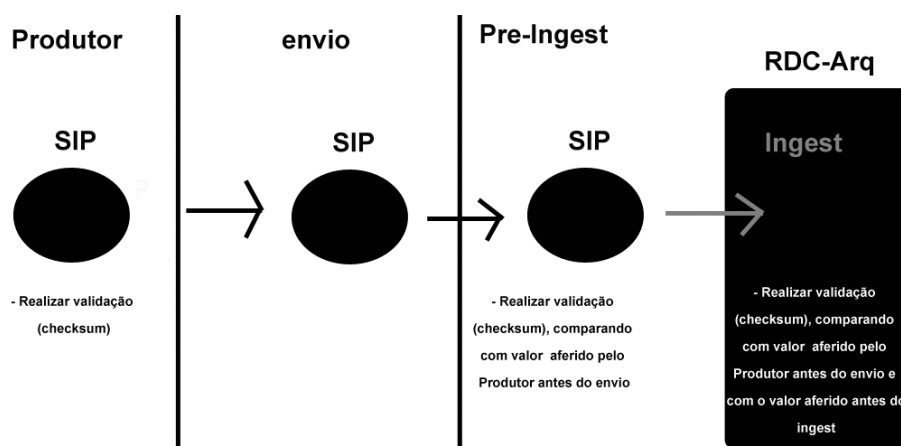
## 6 A VERIFICAÇÃO DE INTEGRIDADE DE UM DOCUMENTO ARQUIVÍSTICO DIGITAL EM UM RDC-ARQ

No intuito de demonstrar-se como é processada a verificação de integridade de um documento arquivístico digital para fins de preservação por longos períodos, realizou-se a simulação de uma transmissão de Pacote de Submissão de Informação (SIP) entre um Produtor e um Repositório Arquivístico Digital Confiável (RDC-Arq).

Um RDC-Arq é um ambiente de preservação e acesso, pelo tempo que for necessário, para documentos arquivísticos digitais, capaz de atender aos procedimentos preconizados pela Arquivologia nas idades corrente, intermediária e permanente, e aos requisitos de um repositório digital (CONARQ, 2015).

Abaixo, na Figura 10, apresenta-se um esboço das sucessivas verificações de integridade de um SIP ao ser enviado para um RDC-Arq, partindo das orientações de IRMT (2016).

Figura 10 - Envio de pacote SIP a um RDC-Arq: verificação de integridade



Fonte: Os autores.

## 7 CONCLUSÃO

Diante das reflexões proporcionadas pela investigação realizada, verificamos que a integridade das cadeias de *bits* dos documentos arquivísticos digitais está relacionada diretamente com a estabilidade das polarizações magnéticas, com a densidade de grãos magnéticos e trilhas do suporte magnético e, por fim, com a relação entre sinal e ruído produzido na leitura dos *bits*, que está limitado pela tecnologia atual. As interações físicas, em níveis quânticos, ocorrem em consequência da busca de equilíbrio térmico, algo que

não pode ser previsto com exatidão, mas que, inevitavelmente, afetará suportes magnéticos.

Assim, cremos que cabe ao Arquivista, profissional apto a atuar como gestor de RDC-Arq's, apropriar-se de conhecimentos oriundos dos algoritmos criptográficos computacionais e da Física para monitorar o decaimento magnético dos suportes digitais que compõe seu parque tecnológico, a fim de garantir, no longo prazo, o acesso contínuo a documentos arquivísticos íntegros, confiáveis e autênticos.

## REFERÊNCIAS

CALLISTER, William D.; RETHWISCH, David. **Ciência e engenharia de materiais: uma introdução**. Tradução Sergio Murilo Stamile Soares. 9ª edição. Rio de Janeiro: LTC, 2016.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). **Orientação Técnica nº 3: cenários de uso de RDC-Arq em conjunto com o SIGAD**. Rio de Janeiro: Arquivo Nacional, 2015. 8 p.

CORMEN, Thomas H. [et al.] **Algoritmos**; [tradução Arlete Simille Marques]. - Rio de Janeiro: Elsevier, 2012. 1042p.

EVANS, Richard Francis L.; LYBERATOS Andreas; CHANTRELL, W. Roy e RICHTER, Hans Juergen. **Thermally induced error: Density limit for magnetic data storage**. Applied Physics Letters · November 2011.

JIAN-GANG Zhu. **New heights for hard drives**. Materials Today. Review Feature. July/August 2003.

MILLAR, Laura (Ed.). **Module 4: preserving electronic records**. London: IRMT, 2009. 57p.

PAMULA, Danuta; ZIEBINSKI, Adam. **Hardware implementation of the MD5 algorithm**. Silesian University of Technology, Gliwice, Poland. 9th IFAC Workshop on Programmable Devices and Embedded Systems Roznov pod Radhostem, Czech Republic, February 10-12, 2009.

RICHTER, Hans Juergen. **The transition from longitudinal to perpendicular recording**. Journal of Physics D: Applied Physics. Volume 40. Number 9. 2007. R149–R177.

ROGERS, Corinne. **A literature review of authenticity of records in digital systems from 'machine-readable' to records in the cloud**. Acervo, Rio de Janeiro, v. 29, n. 2, p. 16-44, jul./dez. 2016.

TANENBAUM, Andrew S. **Structured Computer Organization**. 5. ed. Pearson, 2006.

YUHUI Deng. **ACM Computing Surveys**. April 2011. Department of Computer Science, Jinan University. Guangzhou, 510632, P. R. China.