

Ontologia para mapeamento da dependência tecnológica de objetos digitais no contexto da curadoria e preservação digital

Eloi Juniti Yamaoka

Resumo

Introdução: As transformações tecnológicas que afetam a geração, tratamento e armazenagem de objetos digitais têm desafiado os profissionais quanto à dependência de *hardware* e de *software* para criar, consumir e monitorar o conteúdo de tais objetos, e garantir acesso a seus conteúdos. O trabalho apresenta os problemas decorrentes da mudança de paradigma da informação fixada em papel para o ambiente digital; os avanços alcançados nas teorias e técnicas de curadoria e preservação de objetos digitais; e a descrição de uma ontologia para sistematizar os conceitos basilares da dependência tecnológica de objetos digitais. **Método:** O estudo tem base bibliográfico documental apresentando conceitos e técnicas desenvolvidos no campo da curadoria e preservação digital. Descreve uma ontologia elaborada com o *software* Protégé. **Resultados:** Apresenta os conceitos de documento, documento digital, objeto digital, preservação e curadoria digital e propõe uma ontologia de apoio ao processo de mapeamento da tecnologia necessária para decodificar os objetos digitais em sinais inteligíveis a humanos. **Conclusões:** Uma adequada conceituação de documento digital e objeto digital permite a elaboração de soluções que visam a preservação digital. A ontologia proposta poderá apoiar as organizações no mapeamento dos riscos e das perdas que podem ocorrer numa substituição ou desativação de componentes tecnológicos em ambientes computacionais. Em trabalhos futuros a ontologia deverá ser aprimorada e validada, buscando-se investigar seu alinhamento a requisitos e padrões já especificados para a preservação digital.

Palavras-chave

Documentos digitais. Objetos digitais. Preservação digital. Curadoria digital. Ontologia.

Introdução

A necessidade de fixar informação em algum tipo de suporte acompanha a história da humanidade. Os primeiros registros de informação - as pinturas rupestres - são datados em mais de 40.000 anos. Da pedra, o suporte da informação migrou para outros objetos como placas de barro, metais, pergaminho e papiro.

A adoção do papel como suporte de informação e a invenção da imprensa no século XV permitiu a reprodução em escala de um objeto que assumiu uma importante função na sociedade: o **documento**. O documento passou a ser uma forma de registrar e acessar a informação e também de comprovação ou de evidência de ações realizadas.

Devido ao valor legal atribuído ao documento, surgiu também a preocupação com as condições de posse e guarda de longo prazo desse objeto físico. Conhecimentos específicos nesta área deram origem à Arquivologia ou à Ciência Arquivística (*Archival Science*). A Arquivologia emergiu da necessidade e do saber prático (*savoir-faire*) acumulado ao longo dos anos (ROUSSEAU; COUTURE, 1998), a partir de dois conceitos principais: o arquivo e o documento arquivístico (THOMASSEN, 2001).

Na década de 1940, teve início o desenvolvimento dos computadores e, após aproximadamente sete décadas da criação, esta tecnologia comprova seu papel de agente transformador na sociedade. Porém, as facilidades proporcionadas pelos computadores e demais tecnologias da

informação e comunicação (TICs) trouxeram como um dos efeitos colaterais o contínuo crescimento da produção de novas informações e, conseqüentemente, desafios à sua gestão e consumo.

As transformações tecnológicas são extremamente rápidas, ainda que os conceitos, processos, e o próprio fazer – na escala do humano – não mudam com a mesma velocidade. O próprio conceito de documento como um objeto material, utilizado há cinco séculos, continua arraigado. Embora estudos – como o de Lyman e Varian (2003) – apontem que as informações criadas atualmente são, em grande parte “nato-digitais”, o tratamento do documento digital ocorre de forma similar ao do documento em papel.

Uma diferença fundamental entre o documento em papel e o documento digital é que o segundo requer a mediação da tecnologia para codificar e decodificar os sinais binários em símbolos compreensíveis. Portanto, há uma dependência de *hardware* e de *software* para criar, consumir e monitorar o conteúdo de um documento digital, o que pode ser entendido como “dependência tecnológica de objetos digitais”.

A evolução da tecnologia proporciona novos recursos e facilidades que é acompanhada, também, de rápida obsolescência. Conseqüentemente, os documentos digitais produzidos em tecnologias anteriores podem ficar inacessíveis ou ter seu acesso dificultado. O desafio atual é desenvolver de técnicas e tecnologias que proporcionem a preservação de longo prazo de documentos digitais com a garantia de acesso aos conteúdos armazenados.

Este trabalho apresenta os problemas decorrentes da mudança de paradigma da informação fixada em papel para o ambiente digital, os avanços alcançados nas teorias e técnicas de curadoria e preservação de objetos digitais, e a descrição de uma ontologia concebida para sistematizar os conceitos basilares da dependência tecnológica de objetos digitais.

O artigo divide-se em cinco partes. A primeira conceitua documento, documento digital, documento digital estático, interativo e virtual, apontando suas características e diferenças. A segunda descreve uma anatomia do objeto digital, com base no trabalho Thibodeau (2002), apresentando conceitos que impactam na busca de soluções para a preservação digital. A terceira apresenta os conceitos e as principais estratégias de preservação digital. A quarta descreve a curadoria digital baseado nos conhecimentos desenvolvidos no Digital Curation Centre (DCC)¹; e a quinta, apresenta uma proposta de ontologia de dependência tecnológica e discute o seu potencial de aplicação.

Documento: do papel ao digital

Devido à ubiquidade e uso intuitivo, o termo documento carece de um conceito claro (PÉDAUQUE, 2003). O Dicionário Brasileiro de Terminologia Arquivística define documento como a “[u]nidade de registro de informações, qualquer que seja o suporte ou formato” (ARQUIVO NACIONAL, 2005, p. 73). É uma definição ampla de documento, na qual ficam explícitos dois elementos inerentes ao documento: o suporte e o formato. Pédaque (2003) observa que as transformações no aspecto material, cognitivo, percepção e uso além da convergência entre a escrita e o audiovisual diferem o documento convencional (em papel) daquele em formato digital. A mudança decorrente da migração do objeto informacional do papel para o digital não é somente uma troca de suporte. É uma mudança que exige o repensar tanto dos processos de trabalho como dos aspectos legais, culturais e de negócio.

Documento digital é definido como: “[d]ocumento codificado em dígitos binários, acessível por meio de sistema computacional” (ARQUIVO NACIONAL, 2005, p. 75), sendo que algumas de suas características são fundamentalmente distintas dos convencionais (em papel):

¹ <http://www.dcc.ac.uk/>.

- a) nos documentos convencionais, geralmente os signos são formados pela fixação da tinta no papel. Nos documentos digitais os signos são cadeias de números binários (0 e 1) que representam os signos;
- b) o suporte dos documentos convencionais é - principalmente - o papel, enquanto que nos digitais as mídias são magnéticas ou óticas;
- c) nos documentos convencionais o conteúdo, a forma e o suporte são inseparáveis formando um único objeto, identificado como “documento”. No documento digital, o suporte e o conteúdo são independentes; e
- d) os documentos convencionais podem ser lidos - em geral - a olho nu. Os documentos digitais requerem a mediação da tecnologia para decodificar os sinais binários em conteúdo interpretável por humanos.

A criação do documento digital é realizada por meio de dispositivos eletrônicos como *scanner*, filmadora, máquina fotográfica, gravador de áudio, telefone celular e o computador. A convergência da tecnologia tem dotado os dispositivos com múltiplas funções tendendo maior aproximação dos dispositivos quanto às suas funções. Assim, o próprio processo de criação e a mídia digital condenam o documento digital a ser escravizado pelo *hardware* e pelo *software* utilizados na sua criação.

Os documentos digitais são codificados em três tipos de dados: **o dado de conteúdo** - que constitui o conteúdo de um documento; **o dado da forma** - que possibilita ao sistema reproduzir o documento na forma correta; e **o dado de composição** - que permite determinar qual é o conteúdo e a forma de cada documento (DURANTI; THIBODEAU, 2006).

Acompanhando a evolução da tecnologia, os documentos digitais sofrem transformações, passando da versão digital do documento convencional estático para documentos interativos e virtuais.

Documentos digitais estáticos

O uso do termo “estático” para documento digital é - de certa forma - paradoxal considerando-se a (relativa) facilidade para sua produção, alteração, copia e transmissão. Os objetos digitais estáticos são aqueles que têm a persistência da sua forma e do seu conteúdo garantidos pela definição de um formato. Desta forma, documentos digitais são estáticos quando não fornecem facilidades para alterar o conteúdo e a sua estrutura é invariante (DURANTI; THIBODEAU, 2006).

Documentos digitais interativos

Segundo Duranti e Thibodeau (2006) documentos digitais interativos são aqueles que permitem variação no conteúdo, na forma, ou em ambos. Podem ser não dinâmicos ou dinâmicos. No primeiro, as regras que governam o conteúdo e a forma não variam e o conteúdo apresentado em qualquer instância é selecionado a partir de dados de um sistema. No documento digital interativo e dinâmico as regras que governam o conteúdo e a forma podem variar em quatro situações a saber:

- a) o conteúdo do documento e/ou sua apresentação pode variar devido inclusão ou alteração dos dados;
- b) o conteúdo do documento varia porque os dados são recebidos de fontes externas não armazenadas no sistema;
- c) o documento é produzido por uma aplicação computacional dinâmica que seleciona diferentes arranjos de regras na sua produção;
- d) o documento é produzido por uma aplicação computacional adaptativa ou evolucionária, na qual as alterações ocorrem de forma autônoma às mudanças nos documentos.

Os conceitos atribuídos aos documentos digitais interativos por Duranti e Thibodeau (2006) são tratados por outros autores, porém com a denominação de documentos virtuais.

Documentos virtuais

Watters (1999) considera documento virtual aquele em que não existe o estado persistente, e no qual algumas - ou todas as instâncias - são geradas em tempo de execução (páginas dinâmicas, *applets* Java ou resultados de aplicações, por exemplo). A autora categoriza os documentos virtuais quanto a sua origem, indicados a seguir:

- a) com o uso de gabarito: o conteúdo é inserido em tempo de execução a exemplo de relatórios que podem ser gerados com o uso de um gabarito padrão, no qual a estrutura tem armazenamento persistente e o conteúdo é trazido do banco de dados;
- b) resultante de processamento: são documentos gerados incluindo resultados computacionais e visualizações derivadas do resultado ou da interação com o usuário;
- c) documentos compostos: que podem ser gerados juntando conteúdos de diferentes fontes e apresentados como uma unidade;
- d) metadados (e a geração de metadados): que são derivados de programas de extração e sumarização na produção de documentos virtuais.

Para Myaeng, Lee e Kang (1999) documento virtual consiste de um concentrador (*hub*) que define sua estrutura e uma folha de estilo que define como ele deve ser mostrado quando instanciado. A definição é decorrente de uma arquitetura de documento proposta pelos autores que permite a criação de um novo documento composto, cujas partes podem estar dispersas geograficamente e que são conectadas com diferentes tipos de hiperligações.

Objeto digital

Objeto digital é um objeto de informação, de qualquer tipo e formato, expressa sob a forma digital (THIBODEAU, 2002). A partir dessa definição, Thibodeau apresenta uma conclusão

com importante impacto na preservação digital. Objetos digitais herdaram as propriedades de três classes (THIBODEAU, 2002):

- a) **objeto físico** – como objeto físico, o objeto digital é simplesmente uma inscrição de sinais em uma mídia. O meio físico determina uma convenção para a gravação de dados com densidade e tamanho de blocos diferentes. A inscrição física é independente do significado e, portanto, o computador não sabe se o objeto contém um documento em linguagem natural ou uma foto, por exemplo. A inscrição física não implica em sintaxe, morfologia ou semântica;
- b) **objeto lógico** – como objeto lógico, o objeto digital é reconhecido e processado por software. No nível lógico, a gramática é independente da inscrição física. Um software aplicativo reconhece o formato do objeto – tipos de dados como os códigos American Standard Code for Information Interchange (ASCII) – e os dados de formatação como, por exemplo, o tipo de fonte, os recuos e os estilos; e
- c) **objeto conceitual** – como objeto conceitual o objeto digital é reconhecido e entendido por uma pessoa ou, em alguns casos, reconhecido e processado por uma aplicação de computador. É o objeto “do mundo real”, enquanto unidade significativa de informações, tal como um livro, um contrato, um mapa ou uma fotografia. O conteúdo e a estrutura de um objeto conceitual devem ser contidos de alguma forma no objeto lógico ou nos objetos que representam o objeto na forma digital. No entanto, o mesmo objeto conceitual pode ser representado em diferentes codificações digitais. Por exemplo, um mesmo documento - gravado e reproduzido no processador de texto MS-Word™ e no Adobe Portable Document Format™ (pdf) – pode manter o mesmo conteúdo, aparência e estrutura. Isto é, têm a

mesma aparência visual. Este exemplo revela dois importantes aspectos dos objetos digitais e que têm implicações na sua preservação. Primeiro, existem diferentes codificações digitais do mesmo objeto conceitual; e, segundo, que diferentes codificações podem preservar as características essenciais do objeto conceitual.

O Guia para preservação do patrimônio digital da UNESCO adotou e adaptou o conceito de Thibodeau, acrescentando uma classe adicional – o grupo de elementos essenciais – que contém a mensagem, o propósito ou as características pelas quais foi decidido preservar o material (NATIONAL LIBRARY OF AUSTRALIA, 2003).

Preservar os objetos físico e lógico não garante a obtenção do objeto conceitual. É necessário preservar uma forma de traduzir o objeto lógico para o conceitual. Essa forma de traduzir é determinada pela estratégia de preservação adotada.

Preservação Digital

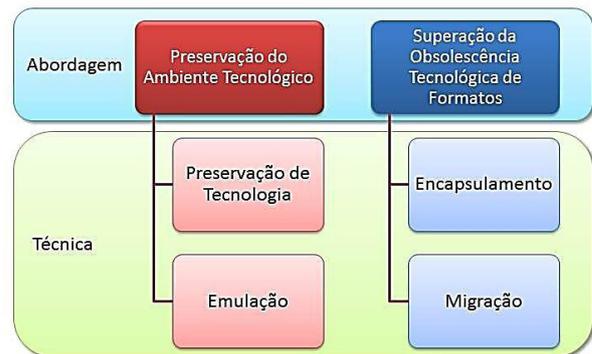
Garantir a continuidade e acessibilidade permanente do patrimônio digital durante o tempo necessário é o objetivo da Preservação Digital a qual busca estratégias (processos, *hardware* e *software*) para a representação do material originalmente apresentado aos usuários (NATIONAL LIBRARY OF AUSTRALIA, 2003). A preservação digital consiste na manutenção da capacidade de apresentar os elementos essenciais dos objetos digitais e proteger este material de eventuais ameaças.

O patrimônio digital sofre dois tipos de ameaças: a física e a técnica. A física é referente aos riscos de danos na mídia e resultantes de incidentes climáticos e a técnica é a dificuldade ou incapacidade de acessar e usar os objetos digitais devido à evolução técnica do *hardware* e do *software* (CONSTANTOPOULOS; DRITSOU, 2007). Instituições de pesquisas e empresas

investem no desenvolvimento de técnicas a fim de reduzir esses riscos.

As técnicas de preservação digital podem ser classificadas em duas principais abordagens: a preservação do ambiente tecnológico; e a superação da obsolescência tecnológica de formatos (Figura 1).

Figura 1 – Técnicas de Preservação Digital



Fonte: adaptado de Lee *et al.* (2002).

A abordagem da preservação do ambiente tecnológico é mais conservadora. Nesta, o ambiente tecnológico original deve ser totalmente preservado para permitir que os objetos digitais gerados possam ser decodificados no futuro. A segunda abordagem busca superar o problema da obsolescência tecnológica dos formatos dos objetos digitais. Um resumo dessas técnicas é apresentado no Quadro 1.

Todas as técnicas apresentadas apresentam vantagens, desvantagens e riscos. A migração é, atualmente, a estratégia mais utilizada (SAYÃO, 2010) e a mais aceita, mas pode acarretar aumento nos custos e a perda de dados; tanto na gravação como na incompatibilidade de formatos (SERRA, 2001).

A preservação de longo prazo de objetos digitais – com a garantia de acesso ao conteúdo – é uma necessidade premente e, nesse contexto, surge o conceito de Curadoria Digital.

Curadoria Digital

A Curadoria Digital é a manutenção e adição de valor a um corpo de informação digital confiável

Quadro 1 – Abordagens e Técnicas de Preservação Digital

Abordagem	Técnica	
Preservação do ambiente tecnológico	Preservação de tecnologia	mantém um museu completo de equipamentos obsoletos para a replicação de qualquer velha configuração. Implica na preservação de software (sistemas operacionais e aplicativos) e hardware. É uma solução de curto prazo, apresenta grau de complexidade e envolve altos custos para manutenção e provimento de espaço físico.
	Emulação	emula plataformas e sistemas operacionais obsoletos em novos sistemas, preservando os programas aplicativos originais. Visa preservar a aparência e as funcionalidades dos objetos digitais. Técnica com defensores e críticos. Requer preservar as especificações detalhadas dos hardwares e sistemas operacionais obsoletos.
Superação da obsolescência tecnológica de formatos	Migração	Estabelece a transferência periódica do objeto digital de uma plataforma tecnológica para outra mais nova, ou de uma geração de tecnologia para uma geração subsequente, com objetivo de preservar a integridade dos objetos digitais e manter a habilidade para que os usuários possam recuperar, visualizar e usar tais objetos.
	Encapsulamento	agrupa e encapsula o objeto digital e todas as informações necessárias para prover o acesso a ele em containers (estruturas lógicas ou físicas).

Fonte: baseado em Lee *et al.* (2002).

para uso corrente e futuro (PENNOCK, 2007). É o ativo envolvimento de profissionais da informação na gestão, incluindo a preservação de dados digitais para uso futuro (YAKEL, 2007). Por Curadoria Digital também se considera a avaliação e a gestão ativa das informações digitais ao longo do seu ciclo de vida, em que se consideram os processos para a manutenção, preservação e agregação de valor aos dados (DIGITAL CURATION CENTRE, [s.d.]).

Um programa de Curadoria Digital tem como objetivo salvaguardar os objetos digitais, possibilitando o acesso e o reuso em todo seu ciclo de vida. O valor despendido para produzir tais objetos tem como retorno o compartilhamento dos dados, o que reduz a duplicação de esforços na criação destes e tornando-os disponíveis para extração de novos conhecimentos. Portanto, a curadoria digital permite:

- manter o documento íntegro e acessível, enquanto este possuir valor jurídico (evidência);
- extrair novos conhecimentos (valor informacional e de pesquisa);
- preservar a memória da sociedade (valor histórico); e

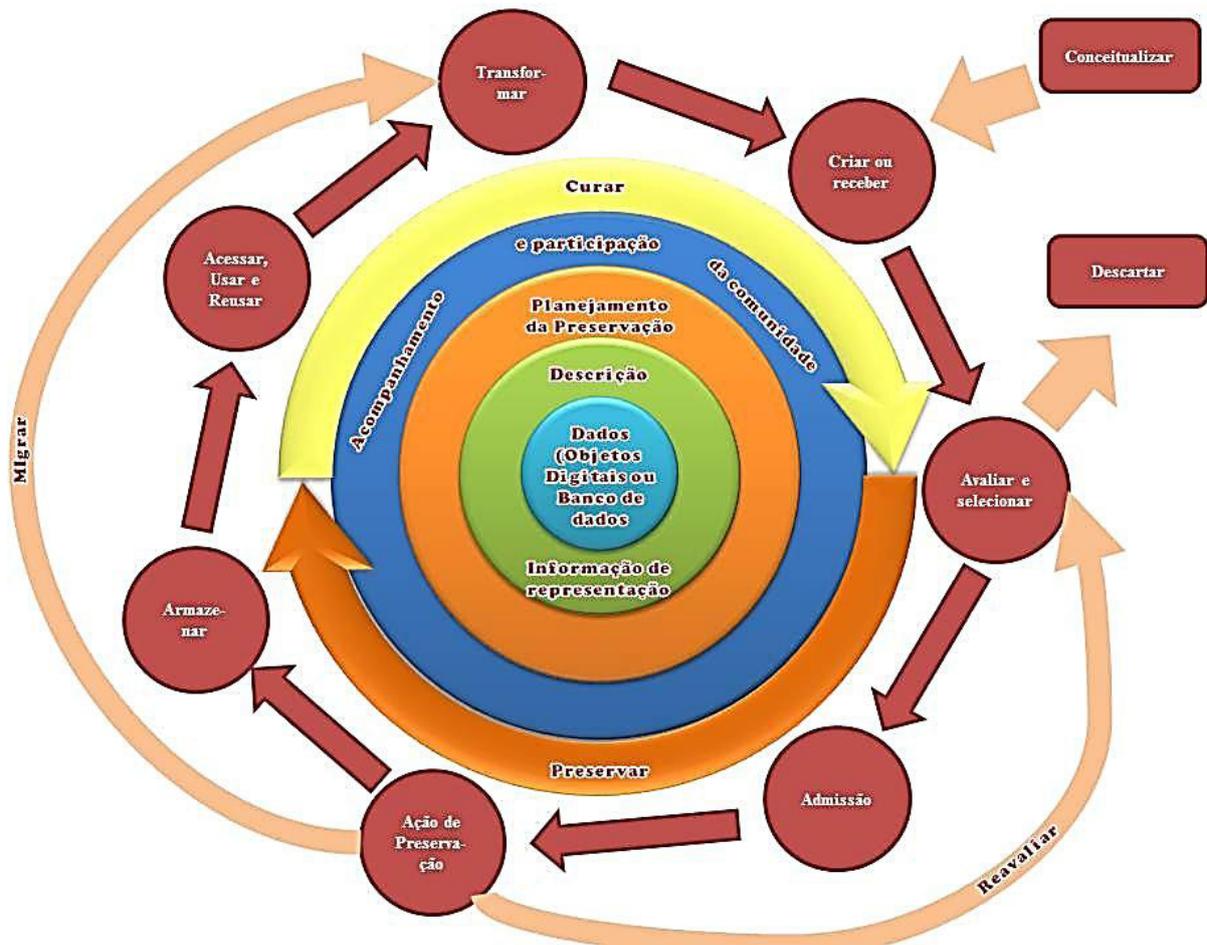
- evitar o retrabalho de recriar os dados já produzidos anteriormente.

Um Modelo de Ciclo de Vida de Curadoria foi desenvolvido pelo Digital Curation Centre e fornece uma visão geral das etapas necessárias para a curadoria e preservação de dados. Esta abordagem de ciclo de vida garante que todas as etapas requeridas sejam identificadas e planejadas, que as ações necessárias sejam encadeadas numa sequência correta, o que garante a manutenção da autenticidade, confiabilidade, integridade e usabilidade do objeto digital (HIGGINS, 2008) (Figura 2).

Os elementos do modelo são detalhados a seguir (HIGGINS, 2008):

- dados:** qualquer informação digital binária. É o centro do ciclo de vida de curadoria e inclui os objetos digitais simples e complexos, e o banco de dados.
- ações:**
 - informação descritiva e de representação:* atribuição dos metadados (administrativo, descritivo, técnico, estrutural e de preservação) para garantir descrição e controle no tempo. Inclui a

Figura 2 – Modelo de ciclo de vida de curadoria digital



Fonte: adaptado de Higgins (2008).

coleta e atribuição da informação de representação requerida para entendimento e “renderização” do objeto digital e dos metadados associados,

- *planejamento da preservação*: é o plano para preservação em todo o ciclo de vida do objeto digital, incluindo a gestão de todas as ações de curadoria,
 - *acompanhamento e participação da comunidade*: é o processo de participação no desenvolvimento de padrões comuns, ferramentas e software, e
 - *curadoria e preservação*: é encaminhar a gestão e as ações planejadas para promover a curadoria e preservação.
- c) **ações sequenciais:**
- *conceitualização*: conceber e planejar a criação de dados, incluindo o

método de captura e opções de armazenamento,

- *criação ou recebimento*: criar dados, incluir metadados administrativos, descritivo, técnico, estrutural e de preservação; ou receber dados dos criadores, das instituições de arquivos, dos repositórios ou dos *data centers*; e, se necessário, atribuir metadados,
- *avaliação e seleção* dos dados para curadoria e preservação de longo prazo. Ser aderente às orientações documentadas, políticas e exigências legais,
- *admissão (ingest)* dos dados para um arquivo, repositório, *data center* ou outra entidade de custódia,
- *ação(ões) de preservação*: buscando garantir que os dados permaneçam autênticos, confiáveis e usáveis, ou

seja, manter sua integridade. As ações incluem limpeza de dados, validação, atribuição de metadados de preservação, informações representação e garantir estruturas de dados ou formatos de arquivos aceitáveis,

- *armazenamento* dos dados de forma segura e aderente aos padrões relevantes,
- *acesso, uso e reuso*, de forma a garantir o acesso aos usuários. Efetivar um robusto controle de acesso e de autenticação,
- *transformação*: criar novos dados a partir do original como, por exemplo, a migração para um formato diferente ou a geração de resultados derivados por seleção ou consulta.

d) **ações ocasionais:**

- *descarte* dos dados que não foram selecionados para curadoria e preservação. Os dados podem ser transferidos para outro arquivo, depósito, data center ou outro centro de custódia. A natureza do dado pode exigir destruição segura,
- *reavaliação e seleção* dos dados retornados por falhas nos procedimentos de validação,
- *migração* do dado para um formato diferente, para garantir a imunidade da obsolescência de *hardware* e *software*.

Além da curadoria digital, a gestão da obsolescência tecnológica é outro complexo desafio.

A gestão da obsolescência tecnológica

Diferentemente do documento em papel, o documento digital exige recursos tecnológicos para decodificar os bits em representações inteligíveis a humanos. A tecnologia é necessária para: permitir o acesso e leitura do objeto digital físico; identificar as regras utilizadas

para a codificação do objeto digital lógico e decodificação para o objeto digital conceitual.

A falta da tecnologia adequada para acesso ao objeto físico ou a decodificação do objeto lógico pode levar a perda total do conteúdo.

A gestão da dependência tecnológica é diretamente afetada com os sucessivos lançamentos de novas versões de dispositivos e *softwares*. A maioria dos produtos e serviços do mercado atual foi criada nos últimos cinco anos e há carência de métodos comprovados que garantam que as informações serão preservadas; que essas informações serão acessadas usando tecnologias disponíveis; e que qualquer informação acessível seja autêntica e confiável (CHEN, 2001).

Um ambiente computacional é constituído de uma quantidade significativa de componentes tecnológicos de inúmeros fornecedores. O conteúdo digital de uma organização é produzido em diferentes *softwares*, que geram objetos com distintos formatos, os quais são armazenados em variados tipos de mídias óticas e magnéticas. Além disso, os objetos digitais podem estar replicados em computadores geograficamente distribuídos. A preservação digital exige a gestão desses componentes, a fim de garantir o acesso ao conteúdo.

Um objeto digital – para ser decodificado – não depende de um ou dois componentes tecnológicos, mas de um **sistema tecnológico**, onde um componente depende de vários outros para a execução da sua função. Esta inter-relação é complexa devido à origem distinta das fontes (internas e externas) de tais componentes.

Visando auxiliar o processo de mapeamento desse ecossistema tecnológico foi concebida uma ontologia que permite a identificação da dependência tecnológica de objetos digitais. Na computação e na ciência da informação, ontologia é um artefato elaborado para modelar o conhecimento de algum domínio ou assunto e permite especificar um vocabulário de asserções que pode ser utilizado por *software*. Segundo

Gruber (2009), uma ontologia define um conjunto de primitivas representacionais, que são as classes, os atributos e os relacionamentos.

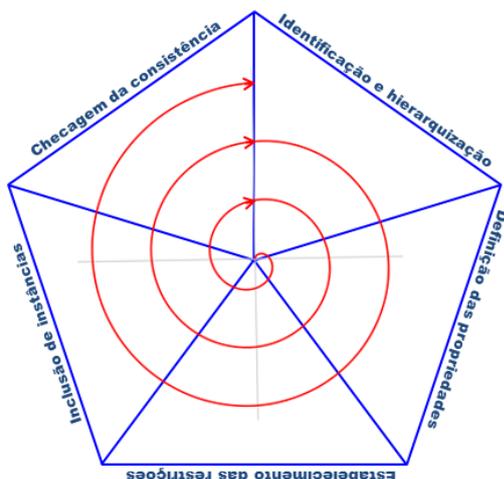
Elaboração de uma ontologia de dependência tecnológica de objetos digitais

A ontologia foi elaborada com o *software* Protégé, na versão Protégé-OWL 3.4.8, desenvolvida no Stanford Center for Biomedical Informatics Research da Stanford University School of Medicine. O Protégé é um conjunto de ferramentas para a construção de modelos de domínio, modelos de conhecimento e sistemas de conhecimento, com ontologias (STANFORD ..., [s.d.]).

A ontologia foi concebida no modelo em espiral (Figura 3), onde em cada ciclo, a ontologia foi aperfeiçoada com ajustes e correções dos conceitos. Cinco atividades completam cada ciclo:

- identificação das classes e a hierarquização;
- identificação e definição das propriedades das classes;
- estabelecimento das restrições;
- inclusão de instâncias; e
- checagem da consistência da ontologia.

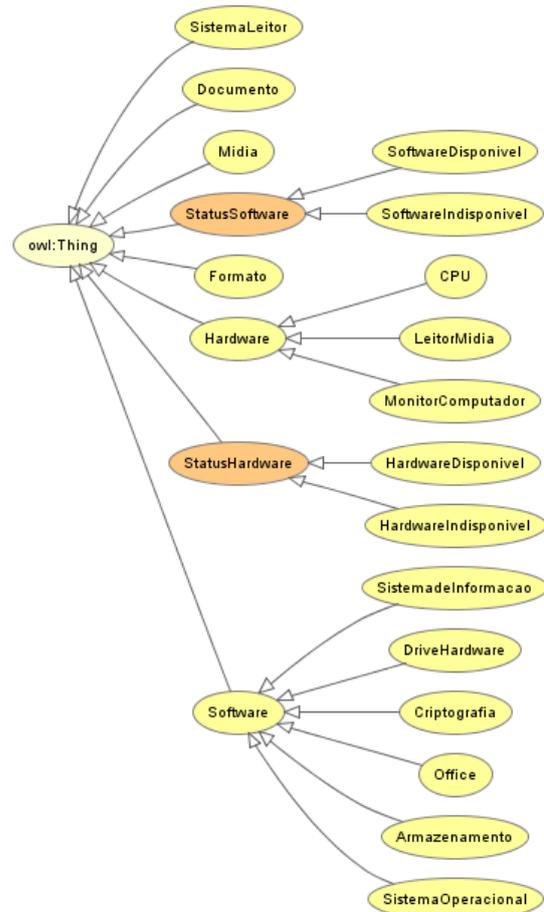
Figura 3 – Espiral de elaboração da ontologia



Fonte: elaborado pelo autor.

A estratégia de elaboração em ciclos possibilita não só correções e aperfeiçoamento em cada ciclo, mas também permite comparar a versão mais recente com as anteriores. Inicialmente, a ontologia foi concebida com a classe documento. Foi a primeira classe a ser criada pela aparente obviedade que - para uma ontologia que visa a preservação do documento digital - esta seria a mais importante (Figura 4).

Figura 4 – Versão inicial da ontologia



Fonte: elaborado pelo autor.

Posteriormente, a classe documento foi substituída por objeto digital no conceito de Thibodeau (2002), que permitiu explicitar as diferenças entre os objetos físicos e lógicos (para máquinas), e o objeto conceitual (para humanos).

A versão mais recente da ontologia apresenta oito classes principais:

- formato de arquivo:** o formato do arquivo é uma das questões centrais em qualquer abordagem de preservação

digital (BARVE, 2007). Entende-se como formato de arquivo a estrutura interna e a codificação de um objeto digital que permite que ele seja processado ou decodificado e apresentado tornando-o compreensível às pessoas (BROWN, 2006). Arquivos com extensão pdf, jpg, tiff, doc, odt, são exemplos de formatos de arquivo. Na ontologia proposta, a classe formato de arquivo representa os formatos utilizados pelos objetos digitais que se pretende monitorar;

- b) o **objeto digital**, no contexto da ontologia, foi modelado seguindo o conceito de Thibodeau (2002), no qual tais objetos herdam propriedades de três classes: **objeto físico**, **objeto lógico** e **objeto conceitual**;
- c) **hardware**: representa todos os dispositivos físicos presentes no ambiente computacional monitorado, como por exemplo, unidade central de processamento, monitor de vídeo, *drive* de CD e DVD, *drive* de disquete etc;
- d) **mídia de armazenamento**: representa os tipos de mídia de armazenamento utilizados no ambiente monitorado, como por exemplo: CD, DVD, disquete 3,5, HD externo, fita tipo/modelo “x” etc;
- e) **software**: representa os *softwares* utilizados no ambiente monitorado, desde o *software* básico até os utilitários e aplicativos; e
- f) **sistema leitor**: concebida visando agregar os componentes necessários para a obtenção do objeto digital conceitual. O sistema leitor mapeia o subsistema necessário para obtenção do objeto conceitual (*hardware*, mídia de armazenamento e formato de arquivo).

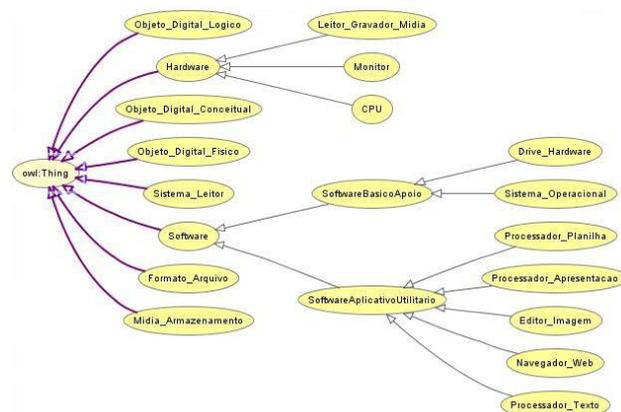
Na concepção da ontologia definiu-se que:

- a) um objeto digital conceitual está representado em um ou mais objetos digitais lógicos;

- b) um objeto digital lógico está armazenado em um (obrigatoriamente) ou mais objetos digitais físicos; e
- c) um objeto digital lógico tem, necessariamente, um formato de arquivo.

A Figura 5 apresenta a ontologia na sua mais recente versão.

Figura 5 – Versão atual da ontologia de dependência tecnológica de objetos digitais



Fonte: elaborado pelo autor.

O Quadro 2 apresenta as classes e as instâncias criadas com o objetivo de verificar a consistência. Os exemplos de instâncias permite também demonstrar a granularidade pretendida na ontologia.

Pretende-se que a versão final da ontologia permita mapear e responder a algumas questões, tais como:

- a) de qual sistema leitor (*hardware* e *software*) um objeto digital é dependente?
- b) de qual(is) *software(s)* um determinado *hardware* é dependente?
- c) de quais *hardwares* e *softwares* uma mídia de armazenamento é dependente?
- d) numa descontinuidade de tecnologia, qual o impacto e quais mídias de armazenamento e formatos de arquivos requerem migração?
- e) qual a dependência do *software* de outro(s)?

Quadro 2 – Exemplos de instâncias

Classe	Instâncias
1. Formato_Arquivo	doc, flv, jpg, ods, odt, pdf, png, wmv xls
2. Hardware	
2.1 CPU	CPU001, CPU 002, CPU 003
2.2 Leitor_Gravador_Midia	LeitorCd001, LeitorCD002, LeitorCartao001, LeitorFitaDat001
2.3 Monitor	Monitor001, Monitor002
3. Midia_Armazenamento	CD, CartaoMS, CartaoSD, DVD, HDEexterno001, PenDrive
4. Objeto_Digital_Conceitu	LivroA, LivroB, Relatorio001, Relatorio002, Artigo001
5. Objeto_Digital_Fisico	OF001, OF002, OF003, OF004, OF005
6. Objeto_Digital_Logico	OL001, OL002, OL003, OL004, OL005,
7. Sistema_Leitor	SistemaLeitor001, SistemaLeitor002, SistemaLeitor003
8. Software	
8.1 SoftwareAplicativoUtilitario	
8.1.1 Editor_Imagem	FotoshopOSxvERSAO001, PAINTsHOPpROuBUNTUv001, fOTOSHOPwINDOWS7x86
8.1.2 Navegador_Web	FirefoxLinuxV001, FirefoxLinuxV002, ChromeWindows7X86V002, ChromeLinuxV002
8.1.3 Processador_Apresentacao	LOImpressV001, LOImpressV002, PP2007, PP2010
8.1.4 Processador_Planilha	Excel2010, LOCalcV003, Lotus123V002
8.1.5 Processador_Texto	LOWriterV002, MSW2003, MSW2010OSX
8.2 SoftwareBasicoApoio	
8.2.1 Drive_Hardware	DriveLeitorCd001, DriveLeitorCD002, DriveLeitorCartao001, LeitorFitaDat001
8.2.2 Sistema_Operacional	Ubuntu10, Ubuntu11, Windows7X86, WindowsXPProfessional

Fonte: elaborado pelo autor.

Desativar ou substituir um componente tecnológico em um ambiente de TIC é sempre uma operação com riscos, podendo inclusive ocasionar perdas de importantes conhecimentos armazenados em objetos digitais. A ontologia apresentada visa mapear e identificar os objetos digitais que serão afetados nessas situações permitindo sua migração para uma tecnologia atualizada.

Considerações finais

Somente o correto entendimento e conceituação um domínio permite o desenvolvimento de uma solução adequada. Este trabalho buscou explicitar e precisar os conceitos de dois objetos que são centrais no tocante à preservação digital: o documento digital e o objeto digital.

O documento digital tem características próprias, entre eles é importante ratificar três características que tem impacto direto nas soluções de preservação: nos documentos digitais o suporte e o conteúdo são independentes; os documentos digitais requerem uma tecnologia de

mediação para apresentá-lo de forma inteligível para humanos; e, os documentos digitais podem ser estáticos, interativos, dinâmicos e virtuais.

O objeto digital pode ser físico, lógico e conceitual. O que se deve preservar é o objeto conceitual, isto é, a forma e o conteúdo do documento digital. Porém, para preservar o objeto conceitual é necessário que os objetos lógicos e físicos sejam mantidos em tecnologias atualizadas.

Gerir o sistema de dependência tecnológica dos objetos digitais é fundamental não só para a preservação dos objetos, mas também para a racionalização do complexo ambiente de TIC, pois permite o mapeamento de quais componentes podem ser desativados, reduzindo custos.

A ontologia apresentada permite reinterpretar a frequente afirmativa de que o documento digital tem dependência de *hardware* e *software*. Os diagramas apresentados nas Figuras 4 e 5 podem auxiliar na identificação de algumas hipóteses como, por exemplo:

- a) que o objeto digital lógico tem relacionamentos com o formato de arquivo, que por sua vez tem relacionamentos com *software* e com o sistema leitor. Portanto, é o formato de arquivo é que tem dependência do *software*; e
- b) que a mídia de armazenamento tem relacionamentos com o *hardware* e com o sistema leitor. Portanto, a mídia de armazenamento é dependente do *hardware*.

Essa diferenciação conceitual também tem impacto na concepção de soluções de preservação. A percepção corrente – de que os documentos digitais são mera substituição do suporte da informação – leva ao raciocínio que as técnicas e os processos aplicados nos documentos em papel podem ser facilmente adaptados aos documentos digitais. No entanto, a preservação digital exige o monitoramento e a atualização da tecnologia para a sincronia entre a tecnologia de mediação e os documentos digitais. Desta forma, o presente trabalho explora os formatos dos objetos digitais (especificações da estrutura e codificação) na tecnologia utilizada (*hardware* e *software*), e nos processos estabelecidos para a curadoria e preservação dos objetos.

Em trabalhos futuros a ontologia deverá ser aprimorada e validada. Deverá também ser investigado se esta proposta se alinha aos requisitos e padrões especificados para a preservação digital, tais como o modelo de requisitos para sistemas informatizados de gestão arquivística de documentos (e-ARQ Brasil, da Câmara Técnica de Documentos Eletrônicos do Conselho Nacional de Arquivos), e o modelo de referência Open Archival Information System (OAIS), que é a norma ISO 14721:2003.

Referências

- ARQUIVO NACIONAL. **Dicionário brasileiro de terminologia arquivística**. Rio de Janeiro: Arquivo Nacional, 2005.
- BARVE, S. File formats in digital preservation. In: INTERNATIONAL CONFERENCE ON DIGITAL LIBRARIES, Bangalore, 2007. **Proceedings...** Bangalore: [s.n.], 2007. Disponível em: <http://drtc.isibang.ac.in:8080/xmlui/bitstream/handle/1849/312/027_p48_sunita-barve_ana-formatted.pdf?sequence=1>. Acesso em: 23 out. 2012.
- BROWN, A. **The PRONOM PUID Scheme**: a scheme of persistent unique identifiers for representation information. London: The National Archives, 2006. (Digital Preservation Technical Paper 2).
- CHEN, S. S. The paradox of digital preservation. **Computer**, v. 34, n. 3, p. 24-28, Mar. 2001. Disponível em: <<http://www.fpdigital.com/resource/files/paradoxofdigitalpreservation.pdf>>. Acesso em: 23 out. 2012.
- CONSTANTOPOULOS, P.; DRITSOU, V. An ontological model for digital preservation. In: INTERNATIONAL SYMPOSIUM IN DIGITAL CURATION, 2007. **Proceedings...** Chapel Hill: [s.n.]. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.129.1241>>. Acesso em: 23 out. 2012.
- DIGITAL CURATION CENTRE. **What is digital curation?** Disponível em: <<http://www.dcc.ac.uk/digital-curation/what-digital-curation>>. Acesso em: 23 out. 2012.
- DURANTI, L.; THIBODEAU, K. The concept of record in interactive, experiential and dynamic environments: the view of interPARES. **Archival Science**, v. 6, n. 1, p. 13-68, Out. 2006. Disponível em: <<http://dx.doi.org/10.1007/s10502-006-9021-7>>. Acesso em: 23 out. 2012.
- GRUBER, T. Ontology. In: LING, L.; ÖZU, T. (Eds.). **Encyclopedia of Database Systems**. Waterloo: Springer-Verlag, 2009.
- HIGGINS, S. The DCC Curation Lifecycle Model. **The International Journal of Digital Curation**, v. 3, n. 1, p. 134-140, 2008. Disponível em: <<http://dx.doi.org/10.2218/ijdc.v3i1.48>>. Acesso em: 23 out. 2012.
- LEE, K.; SLATTERY, O.; LU, R.; TANG, X.; MCCRARY, V. The state of the art and practice in digital preservation. **Journal of Research of the National Institute of Standards and Technology**, v. 107, n. 1, p. 93-106,

2002. Disponível em: <http://ia700705.us.archive.org/23/items/jresv107n1p93/jresv107n1p93_A1b.pdf>. Acesso em: 23 out. 2012.
- LYMAN, P.; VARIAN, H. **How Much Information?** 2003. Berkeley: UC Berkeley. Disponível em: <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/printable_report.pdf>. Acesso em: 23 out. 2012.
- MYAENG, S. H.; LEE, M.-H.; KANG, J.-H. Virtual documents: a new architecture for knowledge management in digital libraries. In: ASIAN DIGITAL LIBRARIES CONFERENCE, 1999. **Proceedings...** Taipei - Taiwan: [s.n.]. Disponível em: <<http://winslab.cnu.ac.kr/Resource/publication/Virtual Document - a New Architecture for Knowledge Management in Digital Libraries.pdf>>. Acesso em: 23 out. 2012.
- NATIONAL LIBRARY OF AUSTRALIA. **Guidelines for the preservation of digital heritage.** Paris: Unesco, 2003. Disponível em: <<http://unesdoc.unesco.org/images/0013/001300/130071e.pdf>>. Acesso em: 23 out. 2012.
- PENNOCK, M. Digital curation: a life-cycle approach to managing and preserving usable digital information. **Library & Archives**, n. 1, Jan. 2007. Disponível em: <http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf>. Acesso em: 23 out. 2012.
- PÉDAUQUE, R. T. **Document: form, sign and medium, as reformulated for electronic documents.** Paris: [s.n.]. Jul. 2003. Disponível em: <http://archivesic.ccsd.cnrs.fr/docs/00/06/22/28/PDF/sic_00000594.pdf>. Acesso em: 23 out. 2012.
- ROUSSEAU, J.-Y.; COUTURE, C. **Os fundamentos da disciplina arquivística.** 1. ed. Lisboa: Dom Quixote, 1998.
- SAYÃO, L. F. Uma outra face dos metadados: informações para a gestão da preservação digital. **Encontros Bibli: revista eletrônica de Biblioteconomia e Ciência da Informação**, v. 15, n. 30, p. 1-31, 2010. Disponível em: <<http://dx.doi.org/10.5007/1518-2924.2010v15n30p1>>. Acesso em: 23 out. 2012.
- SERRA SERRA, J. Gestión de los documentos digitales: estrategias para su conservación. **El Profesional de la Información**, v. 10, n. 9, Set. 2001. Disponível em: <<http://diposit.ub.edu/dspace/bitstream/2445/24347/1/525913.pdf>>. Acesso em: 23 out. 2012.
- STANFORD CENTER FOR BIOMEDICAL INFORMATICS RESEARCH. Protégé Overview. Disponível em: <<http://protege.stanford.edu/overview/>>. Acesso em: 23 out. 2012.
- THIBODEAU, K. Overview of technological approaches to digital preservation and challenges in coming years. In: THE STATE OF DIGITAL PRESERVATION: AN INTERNATIONAL PERSPECTIVE, 2002. **Proceedings...** Washington: CLIR and Library of Congress, 2002. Disponível em: <<http://www.clir.org/pubs/reports/pub107/contents.html/thibodeau.html>>. Acesso em: 22 out. 2012.
- THOMASSEN, T. A first introduction to archival science. **Archival Science**, v. 1, n. 4, p. 373-385, 2001. Disponível em: <<http://dx.doi.org/10.1007/BF02438903>>. Acesso em: 23 out. 2012.
- WATTERS, C. Information retrieval and the virtual document. **Journal of the American Society for Information Science**, v. 50, n. 11, p. 1028-1029, Sep.1999. Disponível em: <[http://dx.doi.org/10.1002/\(SICI\)1097-4571\(1999\)50:11<1028::AID-ASI8>3.0.CO;2-0](http://dx.doi.org/10.1002/(SICI)1097-4571(1999)50:11<1028::AID-ASI8>3.0.CO;2-0)>. Acesso em: 23 out. 2012.
- YAKEL, E. Digital curation. **OCLC Systems & Services**, v. 23, n. 4, p. 335-340, 2007. Disponível em: <<http://dx.doi.org/10.1108/10650750710831466>>. Acesso em: 23 out. 2012.

Ontology for mapping the technological dependence of digital objects in the context of digital curation and preservation

Abstract

Introduction: The technological transformations that affect the generation, treatment and storage of digital objects have challenged the professionals regarding the dependence on hardware and software to create, consume and monitor the content of such objects in order to ensure access to its contents. The paper presents the problems arising from the paradigm shift of information fixed on paper to the digital environment; the progress made in the theories and techniques of curation and preservation of digital objects; and the description of ontology to systematize the basic concepts of technological dependency of objects digital. Method: The study is based on bibliographic and documental research presenting concepts and techniques developed in the field of digital preservation and curation. Describes an ontology developed with the Protégé software. Results: Introduces the concepts of document, digital document, digital object, digital preservation and digital curation and proposes an ontology to support the process of mapping technology to decode the digital objects for signal intelligible to humans. Conclusions: An adequate conceptualization of the digital document and digital object allows the development of solutions aimed at digital preservation. The proposed ontology can assist organizations on mapping risks and losses inherent in a replacement or deactivation of technology components in computer environments. In future work the ontology should be improved and validated, seeking to investigate its alignment to standards and requirements already specified for digital preservation.

Keywords

Digital documents. Digital objects. Digital preservation. Digital Curation. Ontology.

Recebido em 8 de outubro de 2012

Aceito em 30 de outubro de 2012

Sobre o autor:

Eloi Juniti Yamaoka

Bacharel em Administração de Empresas - Fundação de Estudos Sociais do Paraná (FESP), Mestre em Ciência da Informação - Universidade de Brasília, Doutorando em Engenharia do Conhecimento - Universidade Federal de Santa Catarina. Analista - Serviço Federal de Processamento de Dados (SERPRO).

eloijy@yahoo.com.br

Como citar este artigo:

YAMAOKA, E. J. Ontologia para mapeamento da dependência tecnológica de objetos digitais no contexto da curadoria e preservação digital. *AtoZ*, Curitiba, v. 1, n. 2, p. 65-78, jan./dez. 2012. Disponível em: <<http://www.atoz.ufpr.br>>. Acesso em:
