

**UNIVERSIDADE DE SÃO PAULO  
ESCOLA DE COMUNICAÇÕES E ARTES**

**AMARÍLIS MONTAGNOLLI GOMES CORRÊA**

**Preservação digital: autenticidade e integridade de documentos  
em bibliotecas digitais de teses e dissertações**

**SÃO PAULO**

**2010**

**AMARÍLIS MONTAGNOLLI GOMES CORRÊA**

**Preservação digital: autenticidade e integridade de documentos  
em bibliotecas digitais de teses e dissertações**

**Dissertação de mestrado apresentada à  
Escola de Comunicações e Artes da  
Universidade de São Paulo para obtenção  
do título de Mestre em Ciência da  
Informação**

**Área de Concentração: Cultura e  
Informação**

**Orientadora: Profa. Dra. Johanna W. Smit**

**SÃO PAULO**

**2010**

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

CORRÊA, Amarílis Montagnolli Gomes

Preservação digital: autenticidade e integridade de documentos em bibliotecas digitais de teses e dissertações / Amarílis Montagnolli Gomes Corrêa ; orientadora Johanna W. Smit. – São Paulo, 2010.

96 f.

Dissertação (Mestrado) – Universidade de São Paulo, 2010.

1. Preservação digital 2. Bibliotecas digitais 3. Documentos digitais 4. Autenticidade 5. Integridade 6. Repositórios digitais II. Título. III. Título: Autenticidade e integridade de documentos em bibliotecas digitais de teses e dissertações.

Nome: CORRÊA, Amarílis Montagnolli Gomes

Título: Preservação digital: autenticidade e integridade de documentos em bibliotecas digitais de teses e dissertações

Dissertação apresentada à Escola de Comunicações e Artes da Universidade de São Paulo para obtenção do título de Mestre em Ciência da Informação.

Aprovado em:

**Banca examinadora**

Prof. Dr. \_\_\_\_\_ Instituição: \_\_\_\_\_

Julgamento: \_\_\_\_\_ Assinatura: \_\_\_\_\_

Prof. Dr. \_\_\_\_\_ Instituição: \_\_\_\_\_

Julgamento: \_\_\_\_\_ Assinatura: \_\_\_\_\_

Prof. Dr. \_\_\_\_\_ Instituição: \_\_\_\_\_

Julgamento: \_\_\_\_\_ Assinatura: \_\_\_\_\_

Aos meus pais, à minha irmã e  
aos meus queridos avós  
Maria José e Clayton (*in memoriam*).

## **Agradecimentos**

À minha família pelo carinho, apoio e incentivo que me conduzem; à quem devo e dedico as minhas conquistas;

À minha orientadora, Profa. Dra. Johanna, por ter aceito o desafio, pela atenção, paciência e contribuição para a realização deste trabalho;

Aos amigos e a todos que colaboraram para a realização deste trabalho.

## RESUMO

CORRÊA, Amarílis Montagnolli Gomes. **Preservação digital**: autenticidade e integridade de documentos em bibliotecas digitais de teses e dissertações. 2010. 96 f. Dissertação (Mestrado) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2010.

Os avanços tecnológicos impõem um grande desafio para a preservação do patrimônio digital. Além de garantir a longevidade e acessibilidade dos documentos, é necessário mantê-los autênticos e íntegros para que os usuários confiem neles como fonte de informação e objeto de estudo. Instituições guardiãs precisam assumir a preservação de suas coleções digitais e construir repositórios digitais confiáveis. Este trabalho recorre a um estudo exploratório para conhecer as práticas de preservação da autenticidade e integridade dos documentos que compõem os acervos das bibliotecas digitais de teses e dissertações das três universidades estaduais paulistas (Universidade de São Paulo - USP, Universidade Estadual Paulista Júlio Mesquita Filho - Unesp e Universidade Estadual de Campinas - Unicamp). As práticas dessas universidades demonstram a necessidade de melhorias para que elas sejam efetivamente capazes de assumir a responsabilidade de salvaguardar o patrimônio digital.

Palavras-chave: Preservação digital. Bibliotecas digitais. Documentos digitais. Autenticidade. Integridade. Repositórios digitais confiáveis.

## ABSTRACT

CORRÊA, Amarílis Montagnolli Gomes. **Digital preservation**: authenticity and integrity of documents in digital libraries of theses and dissertations. 2010. 96 f. Dissertação (Mestrado) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2010.

Technological innovations raise a set of great challenges to the preservation of digital heritage. Besides assuring the longevity and accessibility of digital documents, it is also important to preserve their authenticity and integrity so one can trust and use digital documents as information source and study objects. Memory organizations must assume an active role in preserving their digital collections and building trusted digital repositories. This dissertation searches through an exploratory study to uncover the practices of the digital libraries of theses and dissertations of the three state universities of São Paulo (Universidade de São Paulo - USP, Universidade Estadual Paulista Júlio de Mesquita Filho - Unesp e Universidade Estadual de Campinas - Unicamp) for preserving the authenticity and integrity of their collections. Their preservation practices show that improvements need to be done so they can in fact take the responsibility of preserving digital heritage.

Keywords: Digital preservation. Digital libraries. Digital documents. Authenticity. Integrity. Trusted digital repositories.



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>9</b>
<b>2</b>	<b>BIBLIOTECAS DIGITAIS</b>	<b>12</b>
2.1	Conceitos de biblioteca digital	13
2.2	Panorama histórico	16
2.3	Panorama histórico da biblioteca digital no Brasil	19
<b>3</b>	<b>PRESERVAÇÃO DIGITAL</b>	<b>23</b>
3.1	Estratégias de preservação digital	25
3.2	Padrões de preservação digital	32
3.3	Preservação digital distribuída	35
<b>4</b>	<b>AUTENTICIDADE E INTEGRIDADE DE DOCUMENTOS DIGITAIS</b>	<b>37</b>
4.1	Definindo os conceitos de autenticidade e integridade	38
4.2	Estratégias para verificar a autenticidade e a integridade de documentos digitais	42
4.3	Elementos para verificar a autenticidade e a integridade de documentos digitais	45
<b>5</b>	<b>REPOSITÓRIOS DIGITAIS CONFIÁVEIS PARA PRESERVAÇÃO DA AUTENTICIDADE E DA INTEGRIDADE DE DOCUMENTOS DIGITAIS</b>	<b>50</b>
5.1	Parâmetros para garantir a confiabilidade de repositórios digitais	51
<b>6</b>	<b>BIBLIOTECAS DIGITAIS DE TESES E DISSERTAÇÕES DAS UNIVERSIDADES ESTADUAIS PAULISTAS</b>	<b>56</b>
6.1	A prática das universidades estaduais paulistas: síntese das respostas ao questionário e algumas considerações	58
<b>7</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>65</b>
	<b>REFERÊNCIAS</b>	<b>67</b>

<b>ANEXO A</b>	78
<b>ANEXO B</b>	81
<b>ANEXO C</b>	86
<b>ANEXO D</b>	91

# 1 INTRODUÇÃO

O principal motivo pelo qual preservamos é garantir a memória; assegurar que teremos registro de acontecimentos dos quais participamos ou que julgamos importantes por alguma razão; perpetuação de experiências vividas, de conhecimentos produzidos e de feitos coletivos ou individuais. A memória de um grupo familiar bem como a de uma nação são essenciais para a caracterização, valorização e orientação de cada membro e do grupo como um todo. Preservamos fragmentos do universo da informação aos quais atribuímos valor de testemunho do pensamento e da ação que se projetaram para além da época e das intenções que os geraram (CABRAL, 2002; CONWAY, 2001).

“As chances de preservação das evidências, provas de nossas vidas, pensamentos e conquistas” começaram com a atuação de escribas e clérigos (CONWAY, 2001, p. 12). Em nossa sociedade, a função de preservar a memória coletiva sempre foi delegada a arquivos, museus e bibliotecas, que reúnem, organizam, protegem e divulgam seu acervo (CONWAY, 2001; PEREIRA; RUTINA, 1999)<sup>1</sup>.

A finalidade da preservação não é somente garantir a longevidade do objeto, mas também proporcionar o acesso a ele, pois não adianta conservar o patrimônio se o grupo ao qual ele pertence não o conhece. No ambiente analógico, “preservação e acesso são atividades relacionadas, porém distintas [...]; no universo digital [descarta-se] toda e qualquer noção dúbia [...]” (CONWAY, 2001, p. 14-15) que distinga essas duas atividades.

No fim do século 20, “tornou-se bastante aguda e generalizada a preocupação com a preservação dos testemunhos históricos e culturais [...] aquilo que chamamos ‘patrimônio’” (CABRAL, 2002, p. 63). A sociedade valoriza muito a conservação de originais como provas desse patrimônio, que, quando preservados, se tornam objetos de estudo para as mais variadas finalidades (ROTHENBERG, 1999).

A constante e cada vez maior preocupação com a preservação do patrimônio agora se volta para o grande patrimônio digital (composto de documentos originalmente criados em

---

<sup>1</sup> Os arquivos são as instituições mais antigas no exercício da guarda de documentos. No Ocidente, datam de 3.000 a 500 a.C., quando, entre sumérios, egípcios, assírios e babilônios, estavam, exclusivamente, a serviço das autoridades como arquivos reais, religiosos e diplomáticos (BELLOTTO, 2002). Apenas no fim do século 18, com a Revolução Francesa, é que os arquivos públicos passam a ser abertos aos cidadãos, mas ainda apenas com uso jurídico-administrativo. Para os historiadores, os arquivos começam a se abrir na segunda metade do século 19 (BELLOTTO, 2002).

formato digital e reproduções digitais<sup>2</sup> de documentos analógicos) que estamos construindo. Para que um processo de preservação seja bem-sucedido, são necessários estudos e a definição de conceitos, critérios e técnicas para a identificação, preservação, avaliação e certificação da autenticidade e da integridade desses documentos. Esses estudos estão sendo conduzidos por profissionais de diversas áreas, e nessa pesquisa estão contextualizados em bibliotecas digitais, as provedoras de acesso a parte do patrimônio criado nesse novo cenário, cumprindo, assim, a exemplo das bibliotecas no ambiente analógico, a função de guardiãs<sup>3</sup> da memória coletiva.

Escolher como objeto de estudo a preservação da autenticidade e integridade de documentos digitais, dentre os pontos relevantes do processo de conservação digital, justifica-se pela necessidade de todo guardião oferecer garantias de que o patrimônio sob sua custódia é autêntico e permanece íntegro no decorrer do tempo. Os usuários de uma biblioteca digital, por exemplo, precisam confiar nos documentos digitais para usá-los sem receio, assim como estão acostumados a trabalhar e a crer nos documentos analógicos (CULLEN, 2000). Documentos digitais costumam ser vistos como suscetíveis a manipulações e alterações imperceptíveis, pelo menos à primeira vista, e por isso há muita desconfiança em relação à sua validade como fonte de informação ou objeto confiável.

A provável indiferença dos usuários quanto à tecnologia que cria os recursos informacionais digitais à sua disposição e a credibilidade que as instituições responsáveis pela preservação de nosso patrimônio têm perante a sociedade podem despertar uma questão sobre a real necessidade de elas explicitarem suas políticas de preservação e declarações sobre a autenticidade e a integridade dos documentos digitais sob sua custódia. Mas, depois de alguma reflexão, penso que, mesmo que os usuários se interessem apenas pelo acesso sem se preocupar com os processos nem com as tecnologias que o garantem e mesmo que eles não consultem as políticas nem os atestados apresentados, estes precisam ser divulgados em cumprimento do dever da instituição de ser transparente em suas ações; até porque, com sua divulgação, poderia ter início um processo de conscientização dos usuários. Essa conscientização sobre a importância de um processo de preservação digital bem realizado

---

<sup>2</sup> Neste trabalho, foi adotado o termo **natodigital** para fazer referência aos documentos originalmente digitais. O conceito de documento original tem força no ambiente analógico, mas no digital causa bastante discussão. Rothenberg (2000) diz que é uma característica dos documentos digitais permitir a criação de cópias idênticas e garantir que nenhum dos exemplares seja menos “original” do que o outro, assim como um lote de livros impressos no qual todos são iguais, cada exemplar é *um* original e não *o* original, não existe distinção entre eles. Para Santos (2005), a distinção entre cópia e original aplica-se quando muda o suporte do documento. Em relação às reproduções digitais de documentos analógicos, o termo adotado é **documento digitalizado**.

<sup>3</sup> Neste trabalho, *guardião* é o termo para designar a pessoa física ou a instituição responsável pela custódia e preservação de patrimônio.

resultaria na difusão dessa preocupação e conseqüentemente na cobrança dos usuários em relação às instituições. Obteríamos uma reação em cadeia muito benéfica para a preservação do patrimônio digital, para o usuário e para o próprio guardião.

Conhecer como as bibliotecas digitais brasileiras têm preservado e certificado para seus usuários os atributos autenticidade e integridade e comparar suas atuações com algumas práticas recomendadas internacionalmente permite identificar as melhorias necessárias para uma preservação digital bem-sucedida. A opção, especificamente, pelo estudo das práticas das bibliotecas digitais de teses e dissertações se deve ao fato de elas serem responsáveis por um importante acervo, por contarem com uma infraestrutura nacional bem organizada e por serem obrigadas, desde 2006, a disponibilizar essa produção acadêmica para consulta na internet.

Para desenvolver a pesquisa, selecionamos conceitos-chave (preservação digital; biblioteca digital; autenticidade e integridade) para discussão e observamos na literatura nacional e internacional sugestões de estratégias de preservação e criação de repositórios digitais confiáveis. As bibliotecas digitais escolhidas são as das três universidades estaduais paulistas – Universidade de São Paulo (USP), Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp) e Universidade Estadual de Campinas (Unicamp).

É importante ressaltar que muito da literatura consultada, principalmente sobre autenticidade e integridade de documentos, critérios para a avaliação e comprovação desses atributos, é de responsabilidade de profissionais atuantes na área da Arquivística. Apesar dos diferentes enfoques e propósitos, em razão da relativa proximidade, já que a área é considerada por alguns especialistas uma das ciências da informação (BELLOTTO, 2002), é possível e recomendável que os estudos realizados na área de Biblioteconomia aproveitem as conquistas da Arquivística, que se apresenta bastante organizada em relação a esses temas e com muita experiência ao tratar o documento como prova incontestável (BELLOTTO, 2002). Também é importante justificar a citação de trabalhos publicados há dez anos ou mais em uma pesquisa cujo tema é relativamente recente. Muitos desses textos são de especialistas que se destacaram nos estudos sobre preservação digital e cujos trabalhos, mesmo “antigos”, continuam sendo utilizados, em reconhecimento ao pioneirismo e porque muitas das questões e dos problemas apontados seguem sem solução concreta.

As traduções foram feitas pelo autor deste trabalho, e foram necessárias adaptações para citar fontes em formato digital que não têm data de publicação e/ou paginação. Citação direta sem indicação de paginação apresenta-se como: (BULLOCK, 1999, *on-line*); e a indireta sem indicação de data de publicação: (LIBRARY OF CONGRESS, *on-line*).

## 2 BIBLIOTECAS DIGITAIS

A vida social do homem talvez tenha começado quando surgiram os primeiros vilarejos, aproximadamente no sétimo milênio a.C. (GONTIJO, 2004). Esses vilarejos eram formados por um grupo muito maior de pessoas do que os que formavam as antigas tribos e isso permitiu mais troca de experiências e também exigiu mais organização social. Esse convívio estimulou a comunicação e, conseqüentemente, a produção de conhecimento (SCHRAMM, 1988<sup>4</sup> apud GONTIJO, 2004). Os sistemas de registro, inventados por volta do nono milênio a.C., utilizados até então para, por exemplo, contabilizar e controlar as colheitas e os rebanhos, não eram mais suficientes. Era necessária uma ferramenta mais eficiente para sistematizar e registrar a grande quantidade de conhecimento que estava sendo criado e que precisava ser acumulado e transmitido (GONTIJO, 2004). “O homem sempre se preocupou em registrar suas idéias” (PEREIRA; RUTINA, 1999, p. 11); a intenção de registrar sempre esteve associada à necessidade de lembrar (GONTIJO, 2004).

Construímos um patrimônio reunindo e preservando objetos, práticas e conhecimentos aos quais atribuímos valor, como evidência, testemunho, fonte de informação, de pensamentos e ações, conquistas que julgamos importantes para além da época e intenções que os geraram (CABRAL, 2002; CONWAY, 2001). A importância do patrimônio enquanto memória sempre foi reconhecida e manipulada conforme as conveniências para determinado grupo ou governante. O ato de preservação sempre esteve submetido ao exercício de poder, mas este não é apenas repressor, é também promotor de memórias e esquecimentos<sup>5</sup> (CHAGAS, 2002).

A sobrevivência da memória registrada está sempre ameaçada pela natureza (desgaste natural dos suportes de registro quando em contato com ar, sol, umidade e terremotos, por exemplo) e por destruições provocadas pelo homem, seja por falta de tratamento adequado para preservação, seja por destruições intencionais. Exemplos de destruição não faltam, mas temos também muitos exemplos de tentativas de preservação.

---

<sup>4</sup> SCHRAMM, Wilbur. **The story of human communication**: cave painting to microchip. New York: Harper & Roe, 1988.

<sup>5</sup> Esquecer e perder não são males absolutos, abrem espaço para o novo e para o criativo (CHAGAS, 2002).

As bibliotecas surgiram quando se tomou consciência da importância do registro escrito<sup>6</sup> para a perpetuação e transmissão de conhecimento. Junto com arquivos e museus, elas têm tradicionalmente a missão de reunir, organizar, preservar e tornar acessível a documentação da atividade humana, a memória coletiva para a geração à qual corresponde e para as gerações seguintes (CONWAY, 2001; GONTIJO, 2004; LEVACOV, 1997; PEREIRA; RUTINA, 1999). As bibliotecas são um lugar-memória<sup>7</sup> que neste novo contexto de produção documental compartilharão com outras instituições a responsabilidade de preservar, principalmente, o patrimônio natodigital que não se restringe às tipologias tradicionalmente preservadas (OWEN, 2007).

As bibliotecas já foram “minerais”, “vegetais”, “animais”, voltaram a ser “vegetais”<sup>8</sup>, porém dividindo espaço com suportes produzidos de outra natureza (MARTINS, W., 2002; PEREIRA; RUTINA, 1999). Até a Renascença, as bibliotecas não estavam à disposição de profanos; elas eram tidas como organismos quase sagrados, ou pelo menos religiosos, acessíveis apenas a quem fazia parte de um grupo igualmente religioso ou sagrado (MARTINS, W., 2002). Hoje, com o advento das bibliotecas digitais, o acesso aos documentos pode romper as barreiras do tempo e do espaço.

## 2.1 Conceitos de biblioteca digital

A palavra biblioteca significa coleção de livros, basicamente organizados pela classificação do conhecimento.

Carrega uma imagem construída ao longo do tempo: o sacrário do saber, que em uma ambiência austera convida ao recolhimento do pensar. [...] A biblioteca tradicional é uma instituição secular de grande importância social. [...] É como ponte que liga o universo de usuários e o vasto mundo das idéias. Tem como função a coleta, organização e disseminação de informação, geral e/ou especializada, no âmbito de uma comunidade específica (SILVA; JAMBEIRO; BARRETO, 2006; p. 265).

---

<sup>6</sup> O registro escrito surge como a alternativa mais eficiente à articulação, ao registro e à comunicação. A escritura permitiu fixar os saberes práticos e teóricos (GOULEMOT, 2006); uma civilização é o que ela escreveu, seja com monumentos, seja com o alfabeto (CANFORA, 2006).

<sup>7</sup> No século 18, o volume cada vez maior de publicações torna as bibliotecas indispensáveis (GOULEMOT, 2006).

<sup>8</sup> Classificação a partir da matéria-prima (argila, papiro e pele de animal) usada para a fabricação de suportes informacionais: tabuleta de argila, papiro, pergaminho, respectivamente. Voltou a ser vegetal por causa do uso do papel como matéria-prima para os livros.

A biblioteca tem como missão “operar serviços e sistemas que facilitem o acesso físico e intelectual – e também o acesso de longo prazo – aos seus estoques informacionais” (SAYÃO, 2008-2009, p. 13). Ainda não há consenso sobre o termo que denomina o novo formato das bibliotecas ou os repositórios que têm sido criados para disponibilizar conteúdos organizados na internet. A biblioteca digital *é* uma biblioteca. Guardadas as devidas proporções e fazendo-se os ajustes necessários nos padrões de gerenciamento e trabalho até agora utilizados, “todos os valores e funções [devem continuar] válidos, [...] tendo como fundamento uma analogia direta com a biblioteca tradicional e com a sua missão” (SAYÃO, 2008-2009, p. 13).

Os termos mais comuns na literatura são biblioteca digital, biblioteca eletrônica e biblioteca virtual. O uso indiscriminado deles dá a impressão de que são apenas palavras diferentes para representar a mesma ideia<sup>9</sup>, ou seja, de que se trataria de sinonímia. Outro motivo para a falta de clareza sobre o que é biblioteca digital se deve à diversidade de atores envolvidos em sua história; à percepção das organizações e dos profissionais envolvidos em sua criação; e às expectativas geradas nos mais diversos segmentos da sociedade (SAYÃO, 2008-2009).

Temos que **biblioteca eletrônica** utiliza equipamentos eletrônicos para seu funcionamento e para a leitura de documentos que compõem seu acervo. O qualificador eletrônico se refere aos equipamentos; às informações acessíveis por intermédio de máquinas e a documentos e serviços que dependem de eletricidade para serem usados: neste caso, o conceito de biblioteca eletrônica engloba o de biblioteca digital (TAMMARO, 2008; TENNANT, 1999); por outro lado, no *Dicionário de Biblioteconomia e Arquivologia*, temos que biblioteca eletrônica é fisicamente identificável, provê acesso não apenas ao seu acervo mas também por rede eletrônica, não possui material impresso e faz parte de uma biblioteca digital (CUNHA, 2008).

**Biblioteca virtual** foi o termo lançado pelo criador da internet para designar uma coleção de documentos ligados em rede e que, por muito tempo, foi adotado para definir a nova forma de biblioteca (TAMMARO, 2008). Essa biblioteca seria o resultado da navegação de um usuário por vários *sites* que se ligam uns aos outros, sem oferecer, necessariamente, nenhuma informação substancial<sup>10</sup>; a biblioteca virtual seria oculta, se revelando a cada

---

<sup>9</sup> Tem-se também a impressão de que a escolha por um desses termos depende da preferência do responsável por tal ação. Cunha (1999, p. 258, grifo nosso) chega a dizer, por exemplo: “a biblioteca digital é também conhecida como biblioteca eletrônica (termo *preferido* dos britânicos) [...]”.

<sup>10</sup> Lemos (1998 apud PEREIRA; RUTINA, 1999) faz analogia desse tipo de “navegação” a dicionários que apresentam a definição de dois verbetes, sendo que uma depende da outra, o que chamou de definições



usuário de maneira única à medida que ele colhe, em pontos distintos, as informações que lhe interessam (LEMOS, 1998<sup>11</sup> apud PEREIRA; RUTINA, 1999). O qualificador virtual se refere a algo que não existe, neste caso uma biblioteca transparente, imaterial (COLLIER et al., 1993<sup>12</sup> apud ROWLEY, 2002; TAMMARO, 2008; TENNANT, 1999;).

As **bibliotecas digitais** surgem como oásis de ordem e de permanência de dados, em oposição à “virtual library [que] é a World Wide Web, [...] poderoso organismo que se nutre de informações e cresce de forma caótica” (TAMMARO, 2008, p. 117).

As bibliotecas digitais são organizações que fornecem os recursos, incluindo uma equipe especializada, para selecionar, estruturar, oferecer acesso intelectual, explicar/ aclarar, distribuir, preservar a integridade e garantir a permanência das coleções digitais para que elas estejam prontamente acessíveis e disponíveis para uma determinada comunidade ou para várias (DIGITAL LIBRARY FEDERATION, 1998, *on-line*).

A biblioteca digital deve reunir documentos de forma organizada, sistematizada e pronta para o uso; deve reunir serviços e pessoas que apoiem o ciclo completo de criação, disseminação, uso e preservação de documentos; combinar as funções das bibliotecas tradicionais com serviços de informação em rede (DUGUID, 1997; HENSHAW, 1994<sup>13</sup> apud PEREIRA; RUTINA, 1999). Esta definição pode ser relacionada ao modelo de biblioteca digital proposto por Ross Atkinson<sup>14</sup> (TAMMARO, 2008), segundo o qual a função principal da biblioteca é servir de apoio à criação de conhecimento, especialmente as bibliotecas de pesquisa e as universitárias, que devem basear suas atuações fundamentalmente nas mudanças do processo de comunicação científica.

Não é fácil estabelecer limites entre a definição dos três conceitos<sup>15</sup>. O de biblioteca digital está vinculado ao de biblioteca eletrônica devido à dependência de equipamentos eletrônicos e ao acesso a materiais que não pertencem ao seu acervo; ao de biblioteca virtual, principalmente quando não está direta e/ou exclusivamente relacionada ao acervo de uma

---

circulares, servindo mais para desorientar e confundir do que para esclarecer dúvidas. Levacov (1997) denomina esse tipo de navegação desorientada de acaso. A referência bibliográfica de Lemos (1998) está na próxima nota.

<sup>11</sup> LEMOS, Antônio Agenor Briquet de. Bibliotecas. In: CAMPELLO, Bernadete Santos; CALDEIRA, Paulo da Terra; MACEDO, Vera Amália Amarante (Org.). **Formas e expressões do conhecimento: introdução às fontes de informação**. Belo Horizonte: Escola de Biblioteconomia da UFMG, 1998. p. 345-366.

<sup>12</sup> COLLIER, M. W.; RAMSDEN, A.; WU, Z. The electronic library: virtually a reality? In: HELAL, A. H.; WEISS, J. (Ed.). **Opportunity 2000: understanding and serving users in an electronic library: 15th International Essen Symposium**. Essen: Universitätsbibliothek, 1993, p. 136-146.

<sup>13</sup> HENSHAW, Rod. The library as a place. **College & Research Libraries**, p. 283-285, July 1994.

<sup>14</sup> Bibliotecário da *Cornell University* dedicado ao desenvolvimento de coleções e engajado na campanha pelo acesso aberto, publicação de trabalhos acadêmicos e científicos *on-line*. Disponível em: <<http://www.news.cornell.edu/stories/March06/Atkinson.obit.ws.html>>. Acesso em: 3 jun. 2010.

<sup>15</sup> No próprio *Dicionário de Biblioteconomia e Arquivologia*, são apresentadas muitas características semelhantes na definição dos três tipos de biblioteca sem distingui-los claramente.

biblioteca física já existente. Apesar de biblioteca eletrônica ser um conceito mais abrangente (TENNANT, 1999), “a expressão ‘biblioteca digital’ é empregada mais corretamente e, portanto, é preferível” (TAMMARO, 2008, p. 117)<sup>16</sup>. Neste trabalho foi adotado o termo biblioteca digital e acredita-se que a organização bibliotecária<sup>17</sup> deve inspirar a organização de qualquer coleção, embora algumas vezes os criadores de repositórios digitais não demonstrem vontade de vincular novos serviços a instituições milenares.

## 2.2 Panorama histórico

O termo biblioteca digital é relativamente recente, mas designa a possível concretização de um desejo muito antigo (HARTER, 1997; SAYÃO, 2008), cujo ícone mais remoto e talvez o mais popular seja a Biblioteca de Alexandria: o desejo de reunir todo o conhecimento humano em *um* acervo.

As bibliotecas sempre adotaram as tecnologias correntes para desempenhar suas funções, por isso as digitais podem ser consideradas sua nova fase. As bibliotecas digitais não são uma ameaça às tradicionais nem aos bibliotecários, no entanto exigem muita reflexão sobre a consequência do uso das novas tecnologias para a realização de tarefas básicas (seleção, organização, disseminação); para a interação com os usuários; reflexão sobre como transformar velhos hábitos e como colaborar para o desenvolvimento de novos, pois a sociedade, acostumada com o suporte papel e ambientes analógicos, deverá aprender a usufruir a “conveniência, eficiência e sofisticação das novas tecnologias vinculadas à idéia familiar e confortável de biblioteca” (MCPHERSON, 1997, *on-line*). “O progresso tecnológico mudou a maneira *como* as bibliotecas fazem o seu trabalho, mas não o *porquê* [...]. Pode alterar o conceito fundamental de biblioteca [...] mas não vai modificar substancialmente [a ação de] conectar pessoas à informação” (KUNY; CLEVELAND, 1998, p. 107, grifo nosso). A biblioteca digital precisa herdar as preocupações técnicas e as missões sociais e culturais atribuídas às bibliotecas físicas (LE CROSNIER, 2006) e adaptá-las ao novo cenário.

---

<sup>16</sup> Também segundo Cunha (2008) o termo biblioteca digital é mais apropriado que o termo virtual.

<sup>17</sup> Coleções de conteúdo livres de controle e organização, que sejam a acumulação indiscriminada de documentos, não devem, portanto, ser chamadas de biblioteca digital.

Foi o desenvolvimento da microfotografia que demonstrou ser possível criar um repositório universal e que inspirou algumas “mentes criativas” (SAYÃO, 2008, p. 4) a elaborar e, até mesmo, a tentar concretizar seus protótipos sob forma de projetos.

Um desses projetos<sup>18</sup> foi o de Paul Otlet, fundador da Documentação, que propôs um “meio de tornar disponível para uso todas as fontes escritas ou gráficas de conhecimento” (OTLET, 1934<sup>19</sup> apud RAYWARD, 1997, p. 299), pela extração e organização das informações contidas em qualquer tipo de documento, ou seja, qualquer objeto que servisse como fonte de informação (BUCKLAND, 1998). Seu objetivo era criar um repositório mundial disponível a todos, inclusive repositórios especializados (RAYWARD, 1997). Seu método de organização da informação era sustentado pela Classificação Decimal Universal, que ele adaptou baseado na Classificação Decimal de Dewey e que permitia que fossem feitas associações entre informações semelhantes ou complementares registradas em um mesmo arquivo ou em vários arquivos independentemente do tipo de documento (RAYWARD, 1997). O repositório de Otlet seria a *Bibliographia Universalis*, abrigada em um edifício em alguma parte do mundo e poderia ser acessado por telefone.

Dentre outros projetos, destacamos três restritos a uma dimensão teórica (Wells, Bush, Nelson). Herbert George Wells sugeriu a criação do *cérebro do mundo* ou *enciclopédia mundial permanente*, um repositório de bibliografias e documentos completos em que todo o conhecimento existente seria classificado, resumido, analisado e comparado. Toda essa coleção estaria acessível a qualquer pessoa sem barreiras geográficas, em uma sala de projeção e com possibilidade de reprodução do material de interesse. Wells colocou questões que hoje são fundamentais para as bibliotecas digitais: fontes de informação distribuídas, persistência da informação, preservação, integração das fontes de informação e documentos, e a democratização do acesso<sup>20</sup>, além de descrever o cérebro do mundo como um repositório promotor de inovação e criação de conhecimentos desligado dos canais oficiais (SAYÃO, 2008; TAMMARO, 2008).

---

<sup>18</sup> Os projetos apresentados a seguir datam das décadas de 1930 (Otlet e Wells), 1940 (Bush) e 1960 (Nelson).

<sup>19</sup> OTLET, Paul. **Traité de documentation**. Le livre sur le livre: théorie et pratique (1989, reprint). Liège: Centre de Lecture Publique de la Communauté Française. 1934.

<sup>20</sup> Muito se fala em democratização do acesso à informação na era digital, mas esse é um dos aspectos mais distantes de concretização. O que se entende por democratização do acesso? Por exemplo, muito do conhecimento disponível em formato digital e de interesse acadêmico é acessível apenas por intermédio de uma grande instituição, pois é necessário pagar caro por um artigo ou pela assinatura de periódicos e bases de dados. Assiste-se ao surgimento de periódicos científicos cujo conteúdo é gratuito, mas, por enquanto, fora isso temos acesso livre apenas a alguns documentos de domínio público. É importante conciliar os direitos autorais com o direito de acesso. Democratizar vai além de equipar a sociedade com ferramentas tecnológicas!

A proposta de Vannevar Bush foi enunciada no período pós-Segunda Guerra Mundial. Para resolver o problema enfrentado pelos cientistas que não conseguiam acompanhar o ritmo crescente de publicações, devido à precariedade das técnicas empregadas para sua transmissão e revisão, ele sugeriu a criação de uma máquina pessoal (batizada *Memex*) que facilitasse a recuperação e o acesso à informação desejada no momento oportuno (BUSH, 1945; SARACEVIC, 1996). Sua proposta de organização da informação se aproxima do raciocínio humano, permitindo a indexação associativa (associação, relação entre elementos diferentes) para recuperar os documentos pelo conteúdo, além de propor a autogestão no repositório particular. Bush também comentou a importância da criação de bancos de informação<sup>21</sup>, porque é importante que ela não seja apenas registrada, mas que circule, que seja armazenada, preservada e que esteja sempre acessível (BUSH, 1945).

O projeto de Theodor Holm Nelson, o precursor dos fundamentos da *web*<sup>22</sup> (hipertexto e hipermídia), o *Xanandu*, na década de 1960, tinha por objetivo criar uma base de informações onde os próprios autores publicariam seus textos, relacionando-os a outros documentos da base, além de propor uma solução parcial para a questão do direito autoral: nessa base seria permitido copiar qualquer informação, pois uma vez feita a cópia o autor receberia automaticamente um pagamento (SAYÃO, 2008). Nelson destacou questões similares às de Wells, idealizou um repositório que proporcionaria acesso ao conhecimento e um ambiente propício à criação e interação, à comunicação, entretanto sem a mediação de profissionais.

Os quatro protótipos de repositórios inovadores (Otlet, Wells, Bush e Nelson) se assemelham em vários aspectos: no estabelecimento de relação entre documentos por meio de seu conteúdo; permanência da informação; necessidade de um profissional responsável pela representação dos documentos e seu conteúdo; responsabilidade do usuário em representar o conteúdo; intercâmbio entre usuários e acervos; criação de padrões. Relacionado às propostas para a biblioteca digital, Bush preconizou que o usuário gerenciasse o acervo, que selecionasse, provavelmente a partir de um acervo maior, os documentos que seriam inseridos no seu e fizesse a indexação associativa da forma que lhe conviesse<sup>23</sup>.

---

<sup>21</sup> Adotou-se o termo proposto por Le Coadic (2004), que diferencia banco ou base de informação de banco ou base de dados, definindo estes como os sistemas que gerenciam arquivos de dados, representações codificadas (em 0 e 1) de informação que possibilitam o seu processamento eletrônico.

<sup>22</sup> A *web* define as condições tecnológicas que permitem a concretização da biblioteca digital na forma como é concebida atualmente, seu desenvolvimento efetivo, suas bases teóricas e práticas se fundamentam em resultados de pesquisas realizadas na área da computação denominada recuperação da informação (SAYÃO, 2008).

<sup>23</sup> A indexação associativa e a criação de atalhos pelo usuário para facilitar a consulta aos documentos podem ser comparadas à prática da folksonomia. As designações para essa nova abordagem relativa à indexação e

As previsões sobre a *biblioteca do futuro*<sup>24</sup> diziam que ela seria diferente em sua estrutura, nos processos e na forma de acesso (TAMMARO, 2008). Uma parte das previsões julgava a biblioteca digital a concretização da utopia do acesso universal ao conhecimento humano e a outra parte, um instrumento para atender a necessidade de acesso à informação da comunidade científica. Posteriormente, a visão de como deve ser uma biblioteca do futuro foi ampliada: de coleção de documentos digitalizados para suporte ao ciclo completo da produção de conhecimento, ambiente propício à troca de informações e à comunicação em que os usuários possam compartilhar “conhecimento sem aumentar seus esforços e tenham acesso à informação e a fontes *confiáveis* [no turbilhão de informações oferecidas na rede]” (MARQUES, 2008-2009, p. 27, grifo nosso; temáticas também presentes em DUGUID, 1997 e TAMMARO, 2008).

### 2.3 Panorama histórico da biblioteca digital no Brasil

Além do desenvolvimento tecnológico, o acesso à rede mundial de computadores – a internet – é um fator decisivo para o sucesso das bibliotecas digitais. Até a primeira metade da década de 1990, a internet era usada apenas para comunicação científica e tecnológica em rede nacional e com outros países, e era controlada pela Rede Nacional de Pesquisa. Foi em 1995, com a liberação da internet para uso comercial e sua adoção como meio de comunicação, que teve início a era digital no Brasil (CUNHA; MCCARTHY, 2006; MARQUES, 2008-2009).

Embora pareça ser considerável o aumento do número de usuários de internet no país, segundo dados da Pesquisa Nacional por Amostra de Domicílios (Pnad) realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE), em 2005, sobre acesso à internet<sup>25</sup>,

---

organização de recursos na *web* com participação dos usuários são folksonomia, classificação social ou ainda *social tagging* (CATARINO; BAPTISTA, 2007).

<sup>24</sup> Termo cunhado por J. C. R. Licklider para referir-se a uma “biblioteca completamente baseada em computador” (SAYÃO, 2008, p. 9).

<sup>25</sup> A Pnad sobre o acesso à internet e a posse de telefone móvel celular para uso pessoal é resultante do convênio entre o IBGE e o Comitê Gestor da Internet no Brasil (CGI.br), cujo objetivo é ampliar o conhecimento sobre a utilização das tecnologias da informação e das comunicações no país (Disponível em: <<http://www.ibge.gov.br/home/estatistica/populacao/acesoainternet/default.shtm#>>. Acesso em: 15 jul. 2009).

apenas 21% (32,1 milhões)<sup>26</sup> da população brasileira a partir dos 10 anos de idade acessou a internet pelo menos uma vez em casa, no trabalho, em centros públicos de acesso gratuito ou pago, em estabelecimento de ensino ou outro local (IBGE, 2005a). Em julho de 2009, uma pesquisa do Ibope (2009), considerando os mesmos locais de acesso à internet, registrou 64,8 milhões de brasileiros, a partir dos 16 anos. Parece considerável, em quatro anos, um aumento de 100% na quantidade de brasileiros com acesso à internet. Porém, esse resultado se torna irrelevante ao constatarmos que ele representa apenas pouco mais de 35% da população<sup>27</sup>.

É importante que o acesso de qualidade à internet esteja ao alcance da maioria da população em todas as regiões do país para que possamos afirmar que realmente existe inclusão digital. Sabemos que não basta homogeneizar o acesso aos equipamentos e à rede, como já foi comentado; é necessário capacitar os usuários, oferecer serviços e ambientes que estimulem o desenvolvimento de uma nova cultura para que as “bibliotecas digitais alcancem sucesso junto à comunidade brasileira” (CUNHA; MCCARTHY, 2006, p. 27). Aparentemente, ainda é pequena a importância das bibliotecas na rede mesmo para os acessos com fins educacionais, e é a crescente produção de conteúdos de qualidade em português que poderá contribuir para mudar essa situação (KURAMOTO, 2006; PUNTONI, 2008-2009)<sup>28</sup>.

No Brasil, observamos a tendência de criar bibliotecas digitais restritas a coleções de documentos digitalizados. As bibliotecas digitais ainda não cumprem sua função mais relevante, a que as diferenciaria de um simples banco de informações: ser atuante na produção de conhecimento; proporcionar o encontro e a comunicação entre usuários com interesses afins; favorecer a existência de colégios invisíveis. O ideal é que as bibliotecas digitais evoluam do atual estágio de coleção. Não se trata de crítica nem desmerecimento à importante divulgação de acervos especiais, de patrimônios histórico-culturais, mas espera-se que, em especial, as bibliotecas digitais acadêmicas e de centros de pesquisa se transformem em

---

<sup>26</sup> Desse total, a metade dos acessos foi feita em casa. O acesso discado ainda predomina nesse (57%) e na maioria dos outros ambientes de acesso (52%), e a finalidade predominante é educação e aprendizado (71%) (IBGE, 2005b, 2005c).

<sup>27</sup> Em dezembro de 2009, a *Internet World Stats* divulgou que 36,2% da população brasileira tinha acesso à internet e que em dez anos o crescimento dessa taxa foi de apenas 1,3% (Disponível em: <<http://www.internetworldstats.com/stats10.htm#spanish>>. Acesso em: 28 maio 2010). O *link* para esse levantamento está na seção Internet no mundo no site do Centro de Estudos sobre as Tecnologias da Informação e da Comunicação (CETIC.Br) (<http://www.cetic.br/mundo/index.htm>).

<sup>28</sup> Segundo dados da pesquisa sobre o uso das tecnologias da informação e comunicação no Brasil – TIC Domicílios – realizada entre os meses de setembro e novembro de 2009, pelo CETIC.Br, no quesito Atividades desenvolvidas na internet – treinamento e educação, aproximadamente 71% dos 9.747 entrevistados utilizam a internet para fins educacionais. No entanto, a consulta a bibliotecas para saber da disponibilidade de um livro ou de um artigo fica restrita a 23%. (CENTRO DE ESTUDOS SOBRE AS TECNOLOGIAS DA INFORMAÇÃO E DA COMUNICAÇÃO, 2009a; 2009b). Esse índice talvez possa ser explicado pela falta de divulgação dos serviços da biblioteca ou porque as pessoas buscam outras fontes; o relatório da pesquisa não aborda o porquê da não consulta. Um desafio às bibliotecas digitais é que se preparem para conquistar seu espaço entre pessoas (64% dos entrevistados) que utilizam a internet para realizar atividades/pesquisas escolares.

bibliotecas mais ativas no processo de produção de conhecimento, incorporando recursos de informática que facilitem a pesquisa e a consulta aos documentos e proporcionem a comunicação entre os usuários.

Os atuais projetos de biblioteca digital no Brasil atendem a quatro grandes áreas prioritárias para o desenvolvimento do país (ciência e tecnologia; educação; literatura e humanidades; história e política) e estão subordinados a instituições que já atendiam às demandas dos pesquisadores (CUNHA; MCCARTHY, 2006). Ainda não é possível afirmar qual é a melhor estratégia para o bom desenvolvimento das bibliotecas digitais brasileiras: coordenação espontânea ou centralizada (CUNHA; MCCARTHY, 2006)? Mas talvez devamos apostar no desenvolvimento coordenado centralmente nos moldes do trabalho realizado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) para construir a Biblioteca Digital Brasileira. Dos subprojetos, o de interesse especial para este trabalho é o da Biblioteca Digital Brasileira de Teses e Dissertações.

Uma coordenação centralizada serve para orientar as instituições interessadas em criar sua biblioteca ou repositório e serve também para racionalizar informações, padrões e métodos de trabalho para que as instituições não caminhem paralelamente gastando recursos e tempo em etapas que já foram satisfatoriamente consolidadas por outras. Outra vantagem é a possibilidade da integração e cooperação entre os acervos. No caso de uma coordenação centralizada, é importante que a diversidade de acervos seja contemplada nas orientações e padrões disponíveis para criação e manutenção dos repositórios.

No Brasil<sup>29</sup>, a coordenação centralizada ficou sob responsabilidade do Ibict, que lançou em 2001 o projeto Biblioteca Digital Brasileira, dividida em quatro subprojetos: Biblioteca Digital Brasileira de Teses e Dissertações, Diálogo Científico, Catálogo de Anais de Congresso Eletrônico e Integração de Recursos de Informação, para auxiliar o cumprimento de sua missão: “promover e estimular o registro e a disseminação da produção científica e tecnológica brasileira” (KURAMOTO, 2006, p. 290). No desenvolvimento do projeto foram adotados alguns princípios que visam a garantir seu sucesso, tais como o desenvolvimento e a adoção de *software* de fonte aberta para facilitar a distribuição dos pacotes para todos os interessados; adoção de pacote de modelos tecnológicos diversificados para permitir a integração de acervos e a valorização das instituições; e a adoção de padrões e protocolos internacionais que tornam viável a interoperabilidade entre os sistemas de informação (KURAMOTO, 2006). No entanto, até o momento, não foram encontradas

---

<sup>29</sup> No cenário internacional, há diversas instituições e grupos de pesquisa dedicados a desenvolver padrões, modelos e sugestões de melhores práticas para criação, manutenção e preservação de repositórios digitais.

menções sobre formalizações e orientações para a preservação digital dos acervos das instituições parceiras da Biblioteca Brasileira Digital de Teses e Dissertações.

O bom desempenho da Biblioteca Digital Brasileira projetou o país internacionalmente como o quinto<sup>30</sup> com mais repositórios criados aos moldes da *Open Archives Initiative* (OAI)<sup>31</sup>. O Ibict alcançou seus primeiros objetivos e continua traçando planos para cumprir sua missão mais importante: registrar e disseminar a produção científica e tecnológica do Brasil, além de criar condições para conquistas como um catálogo nacional de produção técnico-científica e de publicações seriadas<sup>32</sup> (KURAMOTO, 2006).

Este trabalho adota o termo biblioteca digital para representar a nova etapa das bibliotecas e acredita que a cooperação é uma boa alternativa para o avanço dos projetos de criação e gerenciamento de bibliotecas digitais. Por enquanto, quase todas as bibliotecas digitais ainda são “simples” repositórios, não proporcionam um ambiente de comunicação e estímulo à criação de conhecimentos como se espera. Observando o projeto da Biblioteca Digital Brasileira de Teses e Dissertações, apesar de sua relevância, infelizmente percebemos que, a rigor, ele não poderia ser chamado de “biblioteca”, porque se trata de um diretório<sup>33</sup>, ele ainda está em um estágio primário. Os projetos brasileiros, importantes para a divulgação do patrimônio científico e cultural em formato digital, ainda precisam evoluir e se transformar em bibliotecas digitais de fato, conforme definições apresentadas na literatura sobre suas funções.

---

<sup>30</sup> Em 27 de junho de 2010, à frente do Brasil (71 repositórios de todos os tipos – bases de dados, bancos de centros de pesquisa, bibliotecas de teses) estão Estados Unidos (308), Reino Unido (166), Alemanha (109) e Japão (77) (REGISTRY OF OPEN ACCESS REPOSITORIES, *on-line*). Kuramoto (2006) acredita que a inscrição de todas as instituições de ensino superior que têm uma biblioteca digital de teses aos moldes do OAI-PMH e a crescente quantidade de periódicos eletrônicos publicados aos moldes do Sistema Eletrônico de Editoração de Revistas (SEER) permitirão que o Brasil alcance o terceiro ou segundo lugar no ranking. Desde a última verificação para atualizar os números apresentados aqui, há sete meses, o Brasil não registrou nenhum novo repositório, enquanto outros países registraram a incorporação de dois a 18 repositórios.

<sup>31</sup> Iniciativa que desenvolve e estimula a adoção de padrões que garantam a interoperabilidade e consequentemente uma disseminação mais fácil e eficiente de conteúdo (OPEN ARCHIVES INITIATIVE, *on-line*).

<sup>32</sup> Outra conquista destacada por Kuramoto (2006) diz respeito à maior independência dos pesquisadores brasileiros em relação a repositórios e publicações internacionais e que essa autonomia representará grande economia para o país. Porém, seria possível realizarmos pesquisa apenas com a produção nacional? É possível ser *tão* independente? Com certeza é importante que o Brasil se mostre produtor de conhecimento, mas continuará sendo consumidor. Talvez aqui esteja um dilema, produzir conteúdo em língua portuguesa para facilitar o acesso da comunidade interna ao conhecimento gerado e ao mesmo tempo desejar que essa produção seja reconhecida pela comunidade internacional.

<sup>33</sup> Uma biblioteca digital pode oferecer a seus usuários acesso a documentos que não pertencem ao seu acervo; mas ser um diretório, ou seja, apenas apresentar a referência de um documento e redirecionar o usuário para o *site* da biblioteca guardiã é satisfatório? Mesmo as bibliotecas digitais das universidades estaduais paulistas que serão apresentadas ainda são apenas coleções.



### 3 PRESERVAÇÃO DIGITAL

Antes de continuar as reflexões deste trabalho, é importante esclarecer que patrimônio digital é constituído por

fontes culturais, educacionais, científicas, administrativas, técnicas, legais, médicas e outros tipos de informação criados digitalmente ou convertidos a partir de um suporte analógico. [...] Materiais digitais podem ser textos, base de dados, imagens fixas (fotografia, gravura, pintura) ou em movimento, gravações sonoras, *softwares* e páginas da internet entre muitos outros tipos que não param de ser criados. Geralmente são efêmeros e exigem manutenção para serem preservados para a geração presente e para as futuras (UNESCO, 2003, *on-line*).

Vivemos a era digital, período curto na história da humanidade, porém de grande impacto e transformação, no qual os encantos das novas tecnologias desviam nossa atenção para as proezas que elas nos permitem realizar. A facilidade que temos para criar, modificar, divulgar e armazenar documentos em formato digital nos leva a acreditar que sua preservação também seja simples (REAGAN, 2009).

Um dos dilemas é combinar a capacidade de produzir com a de preservar, pois elas são inversamente proporcionais (CONWAY, 2001; SAYÃO, 2006). Muito da nossa produção intelectual passou a ser natodigital e armazenada, muitas vezes, sem os mínimos cuidados necessários para tentar garantir sua conservação. As expressões *dark age*, *blank spot* e amnésia digital (BRAND; SANDERS, 1999; REAGAN, 2009; SAYÃO, 2006) são as mais usadas para designar a provável perda de memória, o “vazio” que teremos na história da humanidade correspondente ao nosso período, devido à demora para consolidar práticas de preservação digital ou mesmo para começar a tentar preservar.

Preservação digital exige políticas<sup>34</sup>, estratégias e ações para garantir, no decorrer dos anos, acesso a conteúdo válido apesar dos desafios impostos pelos suportes (mídias) e mudanças tecnológicas. Preservação digital se aplica a documentos natodigitais e a documentos digitalizados.

Políticas de preservação digital documentam o compromisso de uma instituição com a conservação de conteúdo digital para uso futuro; especificam os formatos de arquivos a serem mantidos e o nível de preservação a ser providenciado; e documentam o compromisso de organizar

---

<sup>34</sup> Neste trabalho entende-se política como orientação do comportamento a longo prazo, especialmente em situações repetitivas ou permanentes; decisões prévias e padronizadas que delimitam as ações, o que se deve fazer em determinados casos (MAXIMIANO, 2000).

o processo de preservação de acordo com padrões e melhores práticas que garantam uma custódia responsável. Estratégias e ações de preservação digital devem contemplar a criação de conteúdo, sua integridade e preservação (AMERICAN LIBRARY ASSOCIATION, 2007, *on-line*).

Além de escolher as estratégias, é imprescindível considerar aspectos tecnológicos, organizacionais, sociais, legais e econômicos (SAYÃO, 2006), projetá-los a longo prazo para assegurar as condições básicas, a sustentabilidade, para a continuidade do processo de preservação.

Os documentos digitais não sobrevivem por acaso, ao contrário de documentos analógicos que, mesmo milhares de anos depois de sua criação, ainda são legíveis por qualquer pessoa<sup>35</sup>. Para muitos documentos analógicos, a preservação do suporte e sua forma física garantem a conservação de seu conteúdo. Uma fotografia ou uma carta que fique guardada em condições ambientais adversas (calor e/ou umidade, por exemplo), ainda que apresente traços de deterioração, pode ser contemplada, e seu conteúdo é legível. Exemplos famosos são os inúmeros pergaminhos, tabuletas de argila, a pedra de Roseta, cópias produzidas em mosteiros medievais, entre outros registros culturais e artísticos das mais antigas civilizações e de períodos relativamente recentes da nossa história, que sobreviveram ao tempo mesmo sem cuidados especiais.

No ambiente analógico, há distinção entre a ação de preservação e a de acesso. Um processo de preservação pode estar voltado para o conteúdo ou para o objeto em si, ou seja, pode ter por finalidade preservar o objeto para garantir acesso ao conteúdo bem como pode ter por finalidade manter o objeto em bom estado e para isso limitar ou até mesmo proibir sua manipulação ou consulta.

No ambiente digital, por sua vez, não há espaço para essa distinção, pois preservação e acesso estão intimamente relacionados (CONWAY, 2001). Se, no ambiente analógico, a forma física *livro* é sinônimo do conjunto de palavras registrado nela, no ambiente digital, a mídia (equivalente ao suporte analógico) não é. O documento digital é invisível<sup>36</sup>, preservar apenas o suporte não é suficiente. Porém, quando o suporte, o *hardware* e o *software* são devidamente preservados, é impossível impedir o acesso ao documento e a seu conteúdo.

---

<sup>35</sup> Depoimento de Margaret Hedstrom, da Universidade de Michigan (INTO THE FUTURE, 1998). Afirmação válida se não considerarmos os documentos analógicos que dependem de alguma máquina para que a informação registrada nele seja legível.

<sup>36</sup> Documentos digitais são, em sua essência, combinações binárias legíveis pelos humanos apenas por intermédio de uma máquina e um *software* que, quando se tornam obsoletos, impossibilitam o acesso ao documento, é como se ele deixasse de existir. Esse problema também ameaça documentos analógicos que não são legíveis a olho nu.

Preservar um documento digital significa preservá-lo como foi criado: sua formatação, a funcionalidade dos *links*, recursos multimídia que fazem parte dele e não apenas o conteúdo, pois um documento digital é

[...] composto basicamente por três partes: estrutura lógica, conteúdo e estrutura de apresentação. A estrutura lógica especifica a organização [...] para indicar suas diferentes partes (títulos de páginas, capítulos e seções), o conteúdo [...] é formado pela informação contida em cada parte [...], e a estrutura de apresentação é a forma como o documento aparece na página, sua formatação (SANTOS, 2005, p. 41).

Além da preservação do documento, é necessária conservar informações sobre ele e o contexto em que foi criado e no qual está armazenado. A seguir serão apresentadas as estratégias de preservação mais comuns.

### 3.1 Estratégias de preservação digital

O sucesso de um processo de preservação digital depende do envolvimento dos criadores e dos responsáveis pelo armazenamento e/ou divulgação dos documentos<sup>37</sup>. Muitas instituições, além de bibliotecas, arquivos e museus, têm assumido a custódia do patrimônio digital (SAYÃO, 2006), e conseqüentemente há vários grupos preocupados, estudando e propondo alternativas para uma preservação digital bem-sucedida.

Temos muitas soluções, mas não um “corpo de conhecimentos plenamente consolidados” (SAYÃO, 2006, p. 118). Talvez ainda não tenha decorrido tempo suficiente para ser possível comprovar a eficiência das estratégias adotadas<sup>38</sup>. Nosso patrimônio digital é muito diversificado, e cada tipo de documento tem suas especificidades. A maioria das estratégias de preservação busca solucionar o problema da obsolescência tecnológica, enfrentar os desafios causados pela vida curta das mídias, dos *hardwares*, dos *softwares* e dos

---

<sup>37</sup> Em muitos casos, o responsável pelas duas ações é o mesmo. No entanto, pode ser preocupante que a preservação digital dependa apenas de um indivíduo, pois ele pode não ter todos os recursos necessários para garanti-la a longo prazo (OWEN, 2007).

<sup>38</sup> Por outro lado, já passou tempo suficiente para termos certeza de que políticas de preservação cuidadosamente elaboradas precisam ser implantadas com urgência!

formatos<sup>39</sup> (CONSELHO NACIONAL DE ARQUIVOS, 2004; SAYÃO, 2006) que ameaçam a longevidade e o acesso incondicional aos documentos. No entanto, ainda nenhuma estratégia se mostrou completa o suficiente (ROTHENBERG, 1999), geralmente é preciso adotar mais de uma.

É necessário que os guardiões acompanhem as novidades e tendências tecnológicas, que estabeleçam ações de preservação que se antecipem à obsolescência tecnológica (BULLOCK, 1999; HEDSTROM, 1998). Sugere-se transferir “os documentos em formatos potencialmente obsoletos para o seu subsequente, novas mídias, novos sistemas operacionais, e assim por diante”, para garantir o acesso (BULLOCK, 1999, *on-line*). É importante adotar padrões internacionalmente aceitos e verificar como outras instituições estão trabalhando.

A preservação digital, pelo menos por enquanto, não garante a sobrevivência de todos os aspectos de um documento digital, e por isso é necessário tentar determinar quais aspectos são importantes preservar e quais serão perdidos (ROTHENBERG, 2000)<sup>40</sup>. De qualquer maneira, tendo de escolher<sup>41</sup>, quanto mais aspectos forem corretamente conservados, maior o sucesso do plano de preservação (BULLOCK, 1999). Com a preservação digital busca-se (BULLOCK, 1999; TASK FORCE ON ARCHIVING OF DIGITAL INFORMATION, 1996):

- **Estabilizar** o documento, identificá-lo como um “bloco”, um objeto único. Tarefa especialmente difícil no caso de hiperdocumentos, porque como estabelecer seus limites?<sup>42</sup>
- Preservar a **presença física**, a essência do documento digital (as combinações binárias de 0 e 1 que o formam). Vale ressaltar que assegurar a existência dessa sequência não garante o acesso, garante apenas que o documento existe.
- Preservar a **apresentação**, as especificações de formatação e o *layout*. Muitos documentos têm essas especificações separadas do conteúdo, por isso corre-se o risco de, durante um processo de migração, não preservá-lo tal como foi

---

<sup>39</sup> É válido lembrar que a rápida evolução e obsolescência tecnológica se apresentam vantajosas e essenciais na competição pela supremacia em um mercado bastante competitivo (RIVERA DONOSO, 2009; SAYÃO, 2006). Nesse contexto, *hardwares* e *softwares* podem se tornar obsoletos em menos de três anos (BULLOCK, 1999).

<sup>40</sup> Rothenberg (2000) expõe que ainda não conhecemos os documentos digitais o suficiente, e que por isso é difícil escolher alguns aspectos em detrimento de outros.

<sup>41</sup> Os aspectos a serem preservados e os critérios para sua escolha devem constar na política de preservação e é importante que sejam divulgados para os usuários (MACNEIL; MAK, 2007).

<sup>42</sup> A preservação de objetos multimídia ou outro aspecto qualquer relacionado à funcionalidade de um documento constituem grandes desafios. Rothenberg comenta que “uma dificuldade é delimitar um hipertexto e determinar o que significa preservá-lo e como fazer isso. O que fazer com um documento que apresenta *links* para outros dezesseis documentos? Basta preservar esses 16? E os documentos referenciados por cada um destes? A rede de ligações entre documentos que se forma é monstruosa e impõe cautela e reflexão de quem se propõe a preservar documentos digitais” (INTO THE FUTURE, 1998).

criado<sup>43</sup>. A apresentação de um documento pode ser fundamental para a compreensão de seu conteúdo.

- Preservar o **conteúdo**, garantir acesso a ele, pelo menos em sua forma mais simples, sem formatações especiais.
- Preservar a **funcionalidade** de recursos multimídia (vídeos, áudio, imagem, gráficos interativos, ferramentas de busca, *links* etc.), dispositivos dinâmicos, recursos de busca por palavras, ou barras de ferramentas.
- Preservar a **autenticidade** e a **integridade** do documento e garantir ao usuário que não houve alteração (intencional ou não) durante as inúmeras “manipulações” (transferências de formato ou de mídia, por exemplo) pelas quais o documento foi submetido no decorrer do tempo para preservá-lo<sup>44</sup>.
- Garantir a **localização** e a **referência** de um documento para que seja possível localizá-lo com precisão, distinguir suas versões e distingui-lo de outros.
- Preservar a **proveniência**<sup>45</sup>, a origem e o histórico de custódia do documento. Quando essas informações são devidamente preservadas auxiliam na comprovação da autenticidade e integridade.
- Preservar o **contexto**, as informações sobre *hardware* e *software* usados em sua criação e necessários para visualizá-lo; sobre suas ligações com outros documentos; e sobre os modos de distribuição<sup>46</sup>.

Existem muitas sugestões e tentativas de preservar todos esses aspectos dos documentos digitais. Na literatura, as estratégias de preservação mais comentadas são reprodução em suporte analógico; criação de um museu tecnológico; migração; e emulação.

A **reprodução em suporte analógico** sugere reproduzir os documentos digitais em papel ou microfilme, suportes estáveis cuja longevidade é conhecida, podendo ser, em média, de 300 anos (HEDSTROM, 1998). Essa técnica é questionável por descaracterizar completamente os documentos digitais, em especial os que apresentam componentes

---

<sup>43</sup> Migração é uma estratégia de preservação que será apresentada mais adiante. Durante o processo, o documento fica vulnerável à corrupção, pois alguma especificação pode não ser corretamente transferida para a nova mídia, por exemplo.

<sup>44</sup> Esses dois conceitos serão discutidos no capítulo 4, assim como estratégias para sua verificação e preservação. Ambos estão diretamente relacionados, porém, um documento não apresenta, necessariamente, esses dois atributos. Para a preservação da integridade, os aspectos que merecem atenção são: conteúdo, permanência, referência, proveniência e contexto (TASK FORCE ON ARCHIVING OF DIGITAL INFORMATION, 1996).

<sup>45</sup> Conceito originário da Arquivística (BULLOCK, 1999).

<sup>46</sup> Para preservar documentos digitais pode ser necessário anular as dependências tecnológicas e desativar *links*; preservar o contexto de documentos digitais é um grande desafio (BULLOCK, 1999).

multimídia. Poderia funcionar para documentos digitais “simples”, textos que não tenham componentes multimídia nem *links*, e mesmo assim ele será descaracterizado porque deixa de ser digital, que é a sua primeira característica, e perde seus atributos como a possibilidade de cópia perfeita, busca interna por palavras, manipulação e fácil distribuição, entre outros (ROTHENBERG, 1999). De qualquer forma, a reprodução em suporte analógico ainda é a estratégia de preservação mais viável para ambientes em que complexos processos de preservação digital são inviáveis. Devemos lembrar também que quando o guardião tiver sob sua custódia um mesmo documento tanto na forma analógica quanto na digital ambas devem ser devidamente preservadas<sup>47</sup>.

Outra estratégia é a criação de **museus tecnológicos** que manteriam em funcionamento *hardwares* e *softwares*, ameaçados pela obsolescência, para garantir a legibilidade de qualquer documento que também esteja devidamente preservado. Empecilhos para a criação desses museus são o custo, o espaço, o suporte e as garantias técnicas necessárias, como a durabilidade de componentes do computador cuja vida útil é limitada (BULLOCK, 1999; ROTHENBERG, 1999). Esses museus restringiriam o acesso em larga escala devido a limitações geográficas, mas poderiam colaborar para a recuperação de documentos importantes gravados em mídias antigas e que hoje são considerados perdidos<sup>48</sup> (ROTHENBERG, 1999). A amnésia digital se deve a práticas e políticas equivocadas, ou até mesmo à falta delas, e o custo das operações de recuperação, quando possíveis, é muito alto e sem a garantia de recuperação total (SAYÃO, 2006).

Sem considerarmos a reprodução em suporte analógico, dentre as estratégias apresentadas, a **migração** (*migration* ou *refreshing*) talvez seja a mais popular. Trata-se do processo de transferir o documento digital de um suporte que está se tornando obsoleto ou fisicamente deteriorado para um mais novo e/ou convertê-lo para um formato mais atual e/ou ainda de um sistema operacional para outro (BULLOCK, 1999). Resumidamente, migração é copiar, converter ou transformar documentos digitais de uma geração tecnológica (de mídias, de formatos e plataformas) para outra. Esta técnica não satisfaz a preservação de todos os

---

<sup>47</sup> Os projetos de digitalização de um acervo analógico não devem ter por objetivo descartá-lo, substituí-lo pelas reproduções digitais, e, sim, possibilitar o acesso de interessados que se encontram geograficamente distantes e/ou poupar o documento original de constante manuseio sem impedir o acesso ao seu conteúdo. Digitalizar um acervo duplica os desafios, pois torna-se obrigatória a preservação dos dois, do analógico e do digital.

<sup>48</sup> Encontramos muitos relatos de perdas, muitas delas irreversíveis. Um caso muito comentado é a falha no conjunto de dados do censo realizado em 1960 nos Estados Unidos. Eles foram armazenados em fitas que, mais de uma década depois, só podiam ser lidas por duas máquinas no mundo, e nenhuma delas pertencia ao departamento responsável. Mesmo com uma grande operação de recuperação, muitos dados se perderam e dizem que hoje se sabe mais sobre o censo realizado em 1860. No Brasil, temos a perda irreversível de dados colhidos em observações via satélite na década de 1970, que seriam importantes para acompanhar as mudanças ocorridas na bacia amazônica (SAYÃO, 2006; TASK FORCE ON ARCHIVING OF DIGITAL INFORMATION, 1996).

aspectos citados anteriormente, ela só garante a presença física e o conteúdo e pode não ser suficiente para preservar artifícios multimídia complexos (BULLOCK, 1999).

Existem opiniões contrárias à migração. Afirma-se que há muitos anos essa técnica é usada para manter a legibilidade de informações importantes, mas que mesmo assim a cada ciclo os profissionais responsáveis deparam com novos problemas. Além disso, a migração supõe uma solução para cada novo formato e para cada tipo de documento a ser convertido. Por esse motivo, não é possível acumular muito conhecimento de um ciclo para outro, cada um é tão trabalhoso, caro e problemático quanto o anterior (ROTHENBERG, 1999). “Como não podemos prever como as coisas mudarão, não podemos prever o que fazer para manter determinado documento acessível e legível [...] cada tipo ou cada documento terá o seu momento, suas necessidades e suas especificações para migração [...] cada ciclo será uma experiência única” (ROTHENBERG, 1999, p. 15-16).

Quando um documento é submetido sucessivamente ao processo de migração, corremos o risco de perder muita informação ou corrompê-lo (BULLOCK, 1999; ROTHENBERG, 1999); de qualquer forma, em alguns casos é preferível adotá-la a não ter nenhuma outra estratégia (ROTHENBERG, 1999). A migração deve ser realizada em caráter de urgência sempre que depararmos com inovações tecnológicas antes que *softwares* e/ou *hardwares* se tornem inutilizáveis e ameacem a acesso aos documentos, dificultando ou, até mesmo, impedindo futuras conversões (ROTHENBERG, 1999).

A última estratégia bastante comentada é a **emulação** (*emulation*). Conceituada como imitação do funcionamento de um *software* ou *hardware* por outro programa, sua grande vantagem é garantir, além da presença física e do conteúdo, características como *layout* e funcionalidades originais do documento (BULLOCK, 1999). Essa técnica é considerada a única alternativa confiável para recriar o comportamento original de um documento que exige alta complexidade técnica, uma vez que ele deverá funcionar em computadores cujo comportamento não é conhecido. É necessário desenvolver técnicas que possibilitem encapsular, ou seja, reunir em um *kit*:

- O documento (sequência original de bits);
- O *software* usado para sua criação e/ou reprodução (sequência original de bits executáveis, sistema operacional e outros dados relacionados ao ambiente de funcionamento do *software*);
- Informações que possibilitem a criação do emulador;
- Informações sobre a plataforma de *hardware*; metadados e anotações sobre o histórico do documento, dependências de *software* e *hardware*.

O propósito dessas informações do *kit* é explicar, para uma pessoa que tenha acesso a ele no futuro, como ativar e usar os conteúdos encapsulados, além disso, é importante que as anotações sejam legíveis por humanos (ROTHENBERG, 1999).

Outras estratégias sugeridas são (RIVERA DONOSO, 2009)<sup>49</sup>:

- Produzir **cópia de segurança**, cópias idênticas de todos os documentos de um acervo, inclusive dos considerados de baixo valor, e armazená-las, de preferência em sistemas e locais geograficamente distintos para evitar perda total e alterações em todos os exemplares<sup>50</sup>.
- Adotar **mídias duráveis** para armazenagem, mídias fabricadas com materiais cuja resistência à deterioração é maior, por exemplo, CDs e DVDs de ouro. Dessa forma, não é necessário transferir o documento de uma mídia para outra com muita frequência<sup>51</sup>.
- O **Computador Virtual Universal** é uma estratégia mais teórica do que prática. Semelhante à emulação, porém diferente por não depender de *software* nem de *hardware* para executar a arquitetura básica de programas e sistemas operacionais. Seria um repositório que, quando necessário, possibilitaria a reprodução de um documento em sua constituição original.

O próximo capítulo é dedicado à autenticidade e à integridade, definição de conceitos, estratégias e elementos para verificação desses atributos nos documentos. Mas cabe neste capítulo, dedicado às estratégias de preservação digital, apresentar uma específica para a manutenção dos aspectos de localização, referência e vínculo com outros documentos, importantes para a preservação da integridade, como mencionamos na nota 44: adotar identificadores persistentes

A estabilidade dos *links* depende do uso de identificadores únicos e que continuem ativos pelo tempo que for necessário mesmo que o responsável por sua atribuição deixe de existir; que o documento perca valor na opinião de seu criador; ou que seja transferido de repositório, pois para alguns usuários o documento pode ter valor permanente (SAYÃO,

---

<sup>49</sup> É interessante notar que mesmo dez anos depois não há muita novidade em relação às sugestões de estratégias de preservação. Por isso, textos “antigos” ainda são referência como explicamos na introdução.

<sup>50</sup> Sugere-se armazenar as cópias em pelo menos um local diferente de onde fica o acervo, cópias *off-site*, e distante para evitar que catástrofes como enchente e terremoto não ameacem todas as cópias.

<sup>51</sup> Simples *backup*, o armazenamento das cópias de segurança em suportes estáticos como CDs não é recomendado por uma iniciativa internacional, parceria entre a *Networked Digital Library of Theses and Dissertations* e a *MetaArchive Cooperative* que será apresentada na seção 3.3 (MCMILLAN; SKINNER, 2009).



2007). Atualmente, o meio mais comum para referenciar um documento disponível na internet é citar seu *Uniform Resource Locator* (URL), ou seja, seu endereço. Porém, este é vulnerável à alteração do nome do arquivo, localização ou mídia de armazenamento (SAYÃO, 2007).

Pode-se escolher entre diferentes técnicas para garantir a persistência dos identificadores (SAYÃO, 2007):

- **Redirecionamento:** estratégia mínima que reencaminha os usuários para a localização atual do documento.
- **Servidor de *links*** vinculado a um banco de dados que registra a localização atual de um documento. O *software* servidor *Persistent URL* (PURL) oferecido pela *Online Computer Library Center* (OCLC) mostra-se uma boa opção, porque busca separar o nome do documento de sua localização para aumentar as chances de seu *link* permanecer ativo. O PURL é mais estável que o simples URL, pois não aponta diretamente para a localização, e por isso é útil para ser usado na referência de registros em catálogos bibliográficos. É um sistema mantido e operado pela OCLC; porém, os servidores são gerenciados por instituições que se manifestam interessadas e comprometidas com a manutenção de esquemas de nomes persistentes.
- Contratação de **sistema de identificação oferecido por terceiros**; por exemplo, pela OCLC, com identificadores baseados em padrões abertos como *Handle System*, *Digital Object Identifier* (DOI), e pelo PURL. O **Handle System** foi concebido para operar em um sistema distribuído de servidores com interoperabilidade global para assinalar, armazenar e administrar identificadores ou nomes persistentes (os *handles*). A grande vantagem desse sistema é tornar a identificação de um documento independente de sua localização, propriedade e de outras informações variáveis. O DOI é uma aplicação do *Handle System* para qualquer recurso ao qual possa ser atribuído direito de propriedade intelectual, e está voltado para a gestão e proteção do *copyright*. Além de identificar, o sistema relaciona, se necessário, informações bibliográficas e comerciais ao documento e pode ser aplicado a qualquer forma de propriedade intelectual e documento. O **CrossRef** é uma associação cooperativa independente formada por editores científicos que serve de base referencial. Se o usuário localizar um DOI que por algum motivo não funcionar

como *link* para o documento que deseje consultar, no site do CrossRef ele pode usá-lo como termo de busca para ser direcionado até o documento<sup>52</sup>.

Seja qual for a estratégia adotada, o objetivo final de qualquer processo de preservação deve ser garantir a integridade de um documento, o acesso a ele e ao seu conteúdo (BULLOCK, 1999; REAGAN, 2009; ROTHENBERG, 1999; TASK FORCE ON ARCHIVING OF DIGITAL INFORMATION, 1996).

### 3.2 Padrões de preservação digital

A adoção de padrões é tão importante quanto a escolha das estratégias, porque torna o processo de preservação digital mais fácil, barato e menos freqüente, devido à redução da quantidade de formatos e à demanda por processos específicos para cada um (SAYÃO, 2006). O ideal é optar pelos padrões mais comuns, pelos que aparentemente são mais estáveis, e, de preferência, pelos padrões abertos. Padrão aberto significa que suas especificações e códigos-fonte estão em domínio público e que seu desenvolvimento é resultado do envolvimento de diversos interessados, além de possibilitar a independência perante empresas monopolistas que em muitos casos não estão preocupadas com preservação (RIVERA DONOSO, 2009; SAYÃO, 2006). As vantagens dos padrões abertos para a preservação digital são:

- Tornar possível a criação de programas para reproduzir documentos em formato que não dependa de *software* proprietário<sup>53</sup>;
- Permitir a compatibilidade e interoperabilidade<sup>54</sup> entre programas;
- Diminuir consideravelmente os custos, pois não há necessidade de pagar a licença cobrada por *softwares* proprietários (RIVERA DONOSO, 2009).

---

<sup>52</sup> Desde que as editoras adotem o DOI, o usuário pode acessar as referências indicadas em um artigo independentemente da concorrência existente entre elas.

<sup>53</sup> *Software*, formato ou programa proprietário é aquele cujo código-fonte não está em domínio público, pertence ao seu criador. Precisa-se de autorização para uso, redistribuição ou modificação (GNU, *on-line*).

<sup>54</sup> A adoção de padrões abertos possibilita a integração entre os repositórios distribuídos, entre “estoques de recursos digitais sediados e mantidos por organizações distintas, que, no entanto, se unificam na perspectiva do usuário [...] pela possibilidade de acessar recursos digitais, independentemente da sua localização geográfica, diretamente através do acionamento de *links* presentes em bases de dados, catálogos, índices e portais [...]” (SAYÃO, 2007, p. 66).

É importante adotar padrões antes da criação dos documentos para evitar procedimentos desnecessários de conversão, “desnecessários” se comparados à importante conversão para preservação. Poupar a exposição do documento a procedimentos de conversão reduz o risco de alterações não intencionais. Para cada tipo de documento deve ser escolhido um formato<sup>55</sup>, a variedade de padrões pode gerar dúvidas, mas a tendência é escolher com base em sua longevidade, funcionalidade e estabilidade (LORIST; MEER, [2001?]; RIVERA DONOSO, 2009). Textos, imagens e documentos codificados são os tipos documentais cujos formatos estão mais definidos, difícil pode ser padronizar o formato de um documento composto de tipos diversos; por exemplo, um documento que contenha texto e animações (BULLOCK, 1999; RIVERA DONOSO, 2009). Trabalhar com padrões proporciona interoperabilidade e consenso interdisciplinar (dado que profissionais de diversas áreas estão envolvidos em pesquisas) sobre conceitos, técnicas, procedimentos e independência tecnológica (LORIST; MEER, [2001?]).

O *Open Archival Information System* (OAIS)<sup>56</sup> é um bom exemplo de modelo para construção de repositórios que almejam a preservação a longo prazo, embora não especifique como fazê-lo. Sua proposta é servir de base conceitual, ponto de partida para outras padronizações; não apresenta um modelo pronto e único de implementação, é apenas um guia, pois cabe à instituição definir as melhores práticas para alcançar o objetivo maior (CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS, 2002; ONLINE COMPUTER LIBRARY CENTER, 2002). Esse modelo tem por função “descrever o ambiente tecnológico do sistema, seus componentes funcionais e informações que servem de suporte para os processos” (ONLINE COMPUTER LIBRARY CENTER, 2002, *on-line*). O modelo OAIS é aplicável a bibliotecas e outras instituições culturais, agências governamentais, e setor privado, à qualquer iniciativa que tenha a responsabilidade de criar, disponibilizar, distribuir ou adquirir, a longo prazo, informação em formato digital.

Entre os padrões para preservação do conteúdo e da apresentação dos documentos, o formato PDF é um dos mais usados. No entanto, por ser um formato proprietário,

---

<sup>55</sup> Os formatos recomendados pela *MetaArchive Cooperative* e pela *Networked Digital Library of Theses and Dissertations* para teses e dissertações digitais, devido à facilidade de migração para padrões subsequentes, são PDF (texto); TIF, JPG, GIF e PNG (imagens); MPG, MOV e QT (vídeos); e WAV, MPG e MP3 (áudio) (MCMILLAN; SKINNER, 2009).

<sup>56</sup> O significado do conceito *aberto*, em relação ao sistema de informação, não se refere ao modo de acesso ao acervo, ou seja, não significa que o conteúdo está livre para consulta, mas se refere à tecnologia utilizada para construir o sistema. O OAIS foi criado pelo *Consultative Committee for Space Data Systems* (CCSDS), da Nasa, e oficializado pela norma ISO 14721:2003. A versão brasileira foi regulamentada pela Associação Brasileira de Normas Técnicas (ABNT) com a norma 15472:2007, com o título *Sistemas Especiais de Dados e Informações – Modelo de Referência para um Sistema Aberto de Arquivamento de Informação* (SAAI).

apresentaremos a linguagem de marcação<sup>57</sup> *Extensible Markup Language* (XML). Este padrão é amplamente usado para permitir o compartilhamento de informações estruturadas entre programas, pessoas, entre máquinas e pessoas conectadas em rede local ou via *web* (WORLD WIDE WEB CONSORTIUM, *on-line*). Outras vantagens são que o próprio criador do documento pode estabelecer as marcas; é legível por humanos; pode ser usada diretamente na internet; permite definir protocolos para a troca de dados; as marcações não são predefinidas, novas podem ser criadas quando necessário, no entanto, para que as marcações sejam devidamente compreendidas, é essencial produzir uma gramática para explicá-las (ALMEIDA, 2002).

Outro padrão importante refere-se aos metadados de preservação. Metadados são informações sobre o documento, são dados estruturados que auxiliam na descrição, identificação, gerenciamento, localização, compreensão<sup>58</sup> e preservação de documentos digitais, além de facilitar a interoperabilidade de repositórios, comunicação de informação (CONSELHO NACIONAL DE ARQUIVOS, 2009; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004). Metadados são informações apresentadas interna ou externamente ao documento, fornecidas pelo próprio criador, para lhes conferir validade, (CULLEN, 2000) e pelo guardião, e também devem ser devidamente preservadas<sup>59</sup>. Para a descrição de conteúdo e contexto tecnológico, existe o padrão de metadados Dublin Core, mais simples que o MARC<sup>60</sup>, padrão com base no qual foi criado. O Dublin Core foi planejado para que os próprios autores, mesmo sem conhecimentos técnicos de catalogação, possam descrever suas produções. Ele é composto de 15 elementos opcionais e repetíveis conforme a complexidade do documento e o grau de detalhamento da descrição: título; autor; palavras-chave; categoria (síntese do assunto por sistemas de classificação formais); descrição (textual do conteúdo); responsável pela publicação (autor corporativo); colaborador (autor secundário); data; tipo de documento; formato; acesso (identificador de material disponível na internet); identificador de recurso; fonte (informação sobre um recurso do qual o documento descrito seja derivado, embora se recomende o uso do elemento *relação*); idioma; relação

---

<sup>57</sup> Linguagem de marcação permite destacar em um documento as partes importantes para sua identificação e as mais significativas de seu conteúdo. Destacar significa, na verdade, sinalizar de forma que o computador reconheça cada uma das partes quando solicitado (ALMEIDA, 2002).

<sup>58</sup> Compreensão é a tradução para o conceito de *understandability* dos documentos digitais. Garantir a compreensão é fornecer informações (metadados) suficientes para que o usuário reconheça, por exemplo, os documentos fornecidos como resultado de busca (ONLINE COMPUTER LIBRARY CENTER, 2002).

<sup>59</sup> No capítulo 4, metadados de preservação são discutidos com mais detalhes.

<sup>60</sup> MARC, *Machine-Readable Cataloguing*, padrão de metadados desenvolvido na década de 1960 pela *Library of Congress*, nos Estados Unidos, e adotado por muitas bibliotecas. Possibilita o compartilhamento de informação bibliográfica evitando duplicação de trabalho no processo de catalogação (LIBRARY OF CONGRESS, *on-line*).

(vínculo com outro documento, por exemplo, versão ou adaptação de uma obra, capítulo ou imagens de uma obra)<sup>61</sup>; cobertura (espacial ou temporal do conteúdo); direito autoral; contato (meios para contato com autores e/ou instituição) (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004; SOUZA; VENDRUSCULO; MELO, 2000).

Garantir a preservação dos metadados é uma das ações-chave para tentar obter sucesso no processo de preservação. Metadados podem estar armazenados com o documento, separado ou das duas formas, sendo esta a opção recomendada pelo menos até que seja convencionalizada qual é a melhor alternativa (BULLOCK, 1999; RIVERA DONOSO, 2009). No caso de manter os metadados em locais diferentes, é importante um controle cuidadoso para não gerar versões diferentes (RIVERA DONOSO, 2009).

### 3.3 Preservação digital distribuída

Antes de continuar e antecipando a contextualização da preservação digital em bibliotecas de teses e dissertações, é pertinente apresentar uma estratégia direcionada para esse patrimônio: a preservação digital distribuída. Essa estratégia foi adotada pela *Networked Digital Library of Theses and Dissertations* (NDTLD), organização internacional sem fins lucrativos<sup>62</sup> “dedicada a promover a adoção, criação, uso, disseminação e preservação de TDEs<sup>63</sup> e reproduções digitais de teses e dissertações impressas” (MCMILLAN; SKINNER, 2009, p. 7), que, em 2008, firmou parceria com a *MetaArchive Cooperative*<sup>64</sup>, que há mais de

<sup>61</sup> A *International Federation of Library Associations and Institutions* (IFLA) propôs um modelo de representação descritiva que permite relacionar as informações apresentadas nos registros bibliográficos dos mais variados tipos de documento às necessidades dos usuários que os utilizam. O objetivo do *Functional Requirements for Bibliographic Records* (FRBR) é orientar a representação de documentos para facilitar a tarefa dos usuários, seu objetivo é privilegiar a relação entre os documentos, possibilitar encontrar em *um* registro bibliográfico todos os vínculos possíveis do documento descrito com outros (derivados ou documentos dos quais se originou) ou com suas versões e edições (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS, [2009]).

<sup>62</sup> Criada em 1997, a NDTLD tem o objetivo de “melhorar o ensino de pós-graduação, aumentar a disponibilidade de pesquisas, fortalecer estudantes e universidades, promover a tecnologia de biblioteca digital e reduzir os custos de armazenamento e gerenciamento de TDEs” (MCMILLAN; SKINNER, 2009, p. 7). Na lista atual das instituições-membros da NDTLD, consta apenas a Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio). Apesar de as três universidades estaduais paulistas declararem ser membros, elas só aparecem em uma lista datada de 2003 com outras universidades brasileiras (Disponível em: <<http://scholar.lib.vt.edu/theses/NDLTD/members/20030120.html>>. Acesso em: 8 jun. 2010).

<sup>63</sup> Teses e Dissertações Eletrônicas (TDEs), tradução de *Electronic Theses and Dissertations* (ETDs).

<sup>64</sup> Associação constituída por instituições sem fins lucrativos acadêmicas ou não (entidades governamentais, arquivos, museus, centros de pesquisa, editoras, bibliotecas, entre outras interessadas) que compartilham o objetivo de preservar documentos acadêmicos e culturais. Sua principal missão é apoiar, promover e estender a

seis anos coordena com sucesso um projeto de preservação digital compartilhado, para oferecer serviços de conservação de teses e dissertações digitais.

A principal característica da preservação digital distribuída é o trabalho colaborativo e a dispersão geográfica de servidores para aumentar a solidez contra falhas e minimizar o risco de perda de dados<sup>65</sup>. Esse modelo de preservação distribuída adota o *software* aberto do projeto LOCKSS<sup>66</sup> e recomenda a participação de, no mínimo, três servidores, embora o ideal seja contar com sete. Além do armazenamento dos acervos das instituições parceiras, a rede conta com um sistema de monitoramento constante para identificar possíveis problemas e, quando necessário, repara os documentos corrompidos. Periodicamente, conforme determinação dos guardiões, os repositórios são revisitados para atualizar o conjunto de documentos a serem preservados. E outra vantagem, segundo a MetaArchive, é o baixo custo financeiro.

Muitas etapas dessa estratégia de preservação são automáticas e algumas dependem do bom desempenho da instituição-membro responsável por gerenciar o funcionamento do seu servidor de preservação<sup>67</sup>, através da análise, por exemplo, de relatórios que recebe e do controle da quantidade de servidores em que seu acervo está armazenado, para identificar insuficiências (EDUCOPIA INSTITUTE, 2009; SKINNER; MEVENKAMP, 2010). Nos textos sobre preservação digital distribuída não é mencionada a necessidade de as instituições participantes serem repositórios confiáveis, mas essa característica seria recomendável. No capítulo 5, há uma discussão sobre repositórios digitais confiáveis.

---

prática da preservação digital distribuída e atuar como catalisadora e guia para outras redes de preservação, além de educar as organizações sobre essa estratégia. A MetaArchive oferece serviços a baixo custo e alto grau de impacto para garantir a preservação e acessibilidade a longo prazo (<sup>MC</sup>MILLAN; SKINNER, 2009; EDUCOPIA INSTITUTE, 2009). A única instituição brasileira membro da MetaArchive é a PUC-Rio.

<sup>65</sup> NDLTD e MetaArchive não recomendam o armazenamento estático, conforme comentário na nota 51.

<sup>66</sup> O LOCKSS (*Lots of Copies Keep Stuff Safe* - <http://lockss.stanford.edu/lockss/Home>) começou a ser desenvolvido pelas bibliotecas da *Stanford University* por volta de 1999. A iniciativa propõe o trabalho de preservação comunitária, uma cadeia de instituições que compartilham seus acervos, similar à existência de um mesmo título de livro impresso em diversas bibliotecas: quanto maior o número de exemplares, maior a chance de o título ser preservado. No site do projeto, constam como membros as instituições brasileiras: Biblioteca Virtual em Saúde, Instituto Nacional de Matemática Pura e Aplicada (Impa) e a Universidade Federal da Paraíba.

<sup>67</sup> O servidor de preservação tem acesso restrito, é diferente do servidor que proporciona aos usuários acesso aos documentos, por isso fala-se que é uma rede de preservação apoiada em *dark archives*.

## 4 AUTENTICIDADE E INTEGRIDADE DE DOCUMENTOS DIGITAIS

Vimos que bibliotecas, arquivos e museus sempre foram considerados repositórios confiáveis do legado cultural e histórico da humanidade. Porém, diante do emprego de novas tecnologias para a criação, o armazenamento e a administração desse legado, há quem possa questionar a capacidade dessas instituições para lidar com os desafios impostos, principalmente pelas rápidas mudanças tecnológicas (MACNEIL; MAK, 2007). Essas instituições precisam assumir posições ativas<sup>68</sup> e mostrar que são capazes de preservar o patrimônio digital, garantir sua longevidade, autenticidade e integridade, assim como conservam o analógico. A preocupação com a preservação da autenticidade e integridade de documentos digitais se justifica pela necessidade de os guardiões garantirem que o patrimônio sob sua custódia é autêntico e permanece íntegro no decorrer do tempo.

O responsável pela criação do documento, o responsável pela preservação e o usuário são igualmente importantes no processo de certificação da autenticidade de um documento (DigiCULT, 2002). Buscar métodos adequados que possibilitem garantir a autenticidade e a integridade de documentos digitais exige o estudo de questões tecnológicas, mas, sobretudo, ou primeiramente, o estudo de questões culturais: por exemplo, “como convencer as pessoas de que devem tomar determinadas atitudes ao criar ou preservar um documento?” (DigiCULT, 2002, p. 17). Para respeitar o usuário, é preciso considerar seu processo de pesquisa, identificação, recuperação e uso dos documentos para saber quais informações devem ser oferecidas para que ele julgue se suas exigências quanto à autenticidade são contempladas pelo repositório (MACNEIL; MAK, 2007). Os pesquisadores se preocupam em identificar a integridade da fonte que lhes interessa; em saber qual dos documentos é o mais completo<sup>69</sup>, o mais confiável dentre os que foram recuperados na busca; em saber qual foi a representação mais cautelosamente feita; e se a fonte é autêntica (BEARMAN; TRANT, 1998). “O que encontrei?”, “Como posso ter certeza de que é o que diz ser?”, e “Posso usar?”

---

<sup>68</sup> Halbert e Skinner (2010) comentam que as organizações culturais não estão preparadas para os desafios tecnológicos e como solução delegam a terceiros o gerenciamento de suas coleções digitais desde a aquisição. Os autores reforçam que esse não seria o comportamento em relação ao patrimônio analógico.

<sup>69</sup> MacNeil e Mak (2007, p. 45) fazem uma abordagem interessante: para documentos digitais “não há [necessariamente] uma ação conclusiva que corresponda à noção de ‘publicação’, indicando o final, uma obra concluída. Ao contrário, a ‘publicação’ de um trabalho digital pode significar o oposto, que agora ela está aberta e pronta para cópia”, para ser usada da forma que melhor convier aos usuários.

são dúvidas básicas dos pesquisadores diante dos inúmeros itens recuperados em uma busca por fontes de informação (BEARMAN; TRANT, 1998, *on-line*).

Talvez essas dúvidas sejam menos frequentes quando o usuário utiliza material oferecido por uma biblioteca ou outra instituição que julgue confiável, pois ele tem à disposição fontes que foram, de alguma forma, previamente avaliadas e consideradas válidas. Os usuários não costumam sentir a necessidade de verificar ou questionar a autenticidade dos documentos analógicos consultados em uma biblioteca: estar no acervo disponível para consulta é atestado de credibilidade<sup>70</sup>, além de muitos documentos terem formas físicas consistentes que apresentam informações suficientes sobre sua produção e seus responsáveis (origem, autoria, data de publicação etc.). Em síntese, a forma estável dos documentos analógicos e a seleção feita pelas bibliotecas conferem a eles uma credibilidade raramente discutida.

Ação semelhante – conferir a credibilidade – poderia ser feita em relação aos documentos digitais, se as bibliotecas divulgassem suas políticas de seleção, avaliação e preservação dos atributos autenticidade e integridade. Conhecendo os critérios e as práticas da instituição, os usuários poderiam avaliar se os consideram suficientes para garantir a confiabilidade do acervo. A importância da divulgação das políticas adotadas pelas bibliotecas perante os usuários será novamente comentada em relação aos critérios e às estratégias para avaliar e preservar a autenticidade e a integridade de documentos digitais.

#### **4.1 Definindo os conceitos de autenticidade e integridade**

O problema de fraude documental tende a ser abordado como se fosse uma novidade que surgiu com os documentos digitais, mas a preocupação com a autenticidade de documentos sempre existiu, apenas ganhou mais atenção devido à vulnerabilidade e à

---

<sup>70</sup> Apesar de importante, as bibliotecas não costumam divulgar ou mesmo formalizar sua política de seleção e aquisição. Os bibliotecários podem não se dar conta de que a análise criteriosa nesse processo demonstra preocupação com a origem e autoria dos documentos e determina sua importância, sua relevância para os usuários e, conseqüentemente, para o acervo. Seleção criteriosa é indispensável também nas bibliotecas digitais. A preocupação com a autenticidade é tão antiga quanto o processo de seleção, se assim podemos dizer, mas talvez não fosse percebida pelos próprios profissionais e pelos usuários, que agora sentem necessidade de saber o grau de confiabilidade do documento exibido na tela.



facilidade com que documentos digitais podem ser alterados sem que, em alguns casos, possível identificar tal modificação<sup>71</sup>.

A comparação entre a suscetibilidade dos documentos digitais e a dos analógicos é muito comum, entretanto é feita sem muita clareza. “Quando se afirma que documentos digitais podem ser mais facilmente alterados, o que está sendo comparado é a facilidade para alterar o *arquivo eletrônico* com a dificuldade em alterar um *texto impresso*” (TANSELLE, 2001, p. 134, grifo nosso). Um documento analógico pode sofrer modificações intencionais ou não em dois momentos: na etapa intermediária de produção e no objeto final. Na etapa intermediária, quando o documento está sendo produzido, se a adulteração for benfeita o objeto final não apresentará nenhuma pista; ao passo que, quando houver alteração no objeto final, ela pode ser identificada conforme a habilidade de quem fizer a alteração, a qualidade do material utilizado tanto para confeccionar o objeto quanto para alterá-lo, e conforme a habilidade do perito (TANSELLE, 2001). Documentos analógicos e digitais podem ser igualmente vulneráveis.

Os conceitos de autenticidade e integridade são abordados de maneira diferente conforme o contexto, a finalidade de seu emprego e as especificações da área do conhecimento que os define (MACNEIL; MAK, 2007; SMITH, 2000a). Ainda não existe consenso nem clareza sobre as definições, provável razão pela qual os conceitos parecem ser empregados aleatoriamente. No entanto, algumas definições apresentam pontos semelhantes e tentaremos alcançar algum esclarecimento.

Temos, por parte da *International Organization for Standardization* (ISO) (2001, p. 7), que “documento autêntico é aquele que podemos provar ser o que parece ser; que foi criado por quem alega tê-lo criado; e que foi criado na data apresentada como data de criação”.

A literatura consultada é autoria de profissionais arquivistas e bibliotecários, e algumas vezes é resultado de trabalho conjunto. Os arquivistas definem **autenticidade** como a “qualidade de um documento quando preenche as formalidades necessárias para que se reconheça sua proveniência, independentemente da veracidade do respectivo conteúdo” (BELLOTTO; CAMARGO, 1996, p. 10); qualidade daquele “que é o que pretende ser e que

---

<sup>71</sup> Na Grécia antiga, por exemplo, para garantir a validade de um documento como evidência, para atestar sua autenticidade, usava-se um *selo* (CAVALLO; CHARTIER, 1998). Esse selo surgiu também com a consciência de autoria por parte de poetas líricos, consciência de que não eram mais os deuses mas os homens que relatavam os fatos. Por exemplo, as primeiras linhas da introdução do livro *História*, de Heródoto, sobre a guerra entre gregos e persas: “Ao escrever sua História, *Heródoto de Halicarnasso* teve em mira evitar que os vestígios das ações praticadas pelos homens se apagassem com o tempo [...], permanecessem ignoradas” (HERÓDOTO, 2001, p. 43, grifo nosso). O autor utiliza o seu nome como marca da autenticidade do relato.

está livre de fraude ou corrupção” (MACNEIL; GILLILAND-SWETLAND, 2005a, p. 27). O julgamento sobre a autenticidade deve se basear em afirmações sobre as origens, completude e integridade do documento, ou seja, autenticidade é a qualidade do documento que se mantém inalterado desde sua criação, que é o que diz ser (BEARMAN; TRANT, 1998)<sup>72</sup>.

Para os bibliotecários, autenticidade é a qualidade daquele documento que pode ser original ou ser fiel a um original, ser incorrupto ou ser de origem conhecida, independentemente de ter sido corrompido ou não (SMITH, 2000a); qualidade daquele que, independentemente do formato (manuscrito, impresso ou digital), tem sua integridade intacta; que se pode comprovar ser o que parece ser ou aceito como sendo o que seu detentor diz que ele é (CULLEN, 2000); característica própria de quem ou o que é o que parece ou o que diz ser (LEVY, 2000). Autenticidade diz respeito a questões pontuais, como data de publicação e autoria, e não à análise do conteúdo, se este é verdadeiro ou confiável (LYNCH, 2000).

Os conceitos de integridade e autenticidade algumas vezes aparecem juntos, quase como sinônimos, e outras vezes o de integridade nem é citado. Segundo a ISO (2001, p. 7), afirmar a **integridade** de um documento “significa que [ele] está completo e inalterado”. A integridade é definida pelos arquivistas como o estado dos documentos “completos e que não sofreram nenhum tipo de corrupção ou alteração não autorizada nem documentada” (CONSELHO NACIONAL DE ARQUIVOS, 2010, p. 172). Refere-se ao estado de completude e solidez do documento; comprovar a integridade de um documento digital não significa, necessariamente, que ele permanece igual a quando foi criado, mas que a mensagem registrada nele continua inalterada e que seus aspectos essenciais não estão corrompidos (MACNEIL; GILLILAND-SWETLAND, 2005a). E para os bibliotecários, a integridade demonstra a não corrupção de um documento, a manutenção da mesma sequência de bits desde que foi criado; sobre ela pode-se fazer as seguintes questões: o documento sofreu alterações desde sua criação? Se sim, essas alterações mudaram sua essência? Caso o documento esteja intacto, as declarações sobre ele são verdadeiras ou falsas? (LYNCH, 2000).

É importante ressaltar que a avaliação para comprovar a autenticidade e a integridade de documentos não tem por objetivo comprovar também a veracidade ou confiabilidade do conteúdo propriamente dito, das informações registradas no documento<sup>73</sup> (LYNCH, 2000). A

---

<sup>72</sup> Bearman e Trant (1998) ressaltam que as discussões sobre autenticidade e integridade não devem ser confundidas com discussões sobre autenticação, que é o controle de acesso, de identificação de usuário, de quem tem permissão para fazer o que com determinado documento.

<sup>73</sup> “Provar a autenticidade [e a integridade] de um documento não o torna mais confiável do que era quando foi criado” (DURANTI, 1995, p. 7). Essa conceituação diverge do entendimento dos cientistas sobre autenticidade e

autenticidade se refere à garantia da confiabilidade e da validade do documento como um todo (autoria, data e local de publicação, por exemplo), e a integridade se refere à garantia de validade do conteúdo, atestando sua fidelidade ao conteúdo original, ou seja, atestando que ele não sofreu alterações ilegais<sup>74</sup>.

É difícil distinguir um conceito do outro, por isso entendemos que autenticidade e integridade são atributos complementares para validar um documento, mas que não coexistem necessariamente<sup>75</sup>. Pode haver um documento autêntico cujo conteúdo foi manipulado, assim como pode existir ter um documento cujo conteúdo está intacto, mas sua autoria, por exemplo, não é válida, ou é duvidosa, pois não foi possível comprová-la.

De qualquer forma, autenticidade e integridade não são atributos inerentes a um documento e sim validados e preservados por seu criador e pelo guardião. São características instáveis; a mínima alteração sofrida pelo documento, mesmo que não intencional, se for identificada, pode comprometer a confiança nele ou no responsável pela preservação. A integridade de um documento digital está em risco, por exemplo, quando ele é transferido de um ambiente tecnológico a outro<sup>76</sup>; quando é armazenado em uma mídia ou computador; ou quando o software utilizado para gerá-lo ou armazená-lo é modificado ou substituído (MACNEIL; GILLILAND-SWETLAND, 2005a).

Além de preocupações com alterações no conteúdo de um documento, há uma questão interessante que pode causar muita discussão: documentos digitais são “extremamente sensíveis” (LEVY, 2000, p. 29), podem ser visualizados de maneiras diferentes no mesmo computador no qual foram criados, caso tenha passado por atualizações tecnológicas, ou em

---

integridade. O conceito-chave para eles é *data quality* fortemente relacionado ao conceito de precisão (*accuracy*) (HACKETT; UNDERWOOD; EPPARD, 2008). Para os cientistas, o conceito de integridade não está diretamente relacionado à não corrupção do conteúdo de um documento, mas à qualidade e precisão da coleta de dados. Eles se preocupam com a metodologia utilizada para coleta, testes e geração de dados; eles confiam em dados produzidos conforme padrões rígidos e cujos procedimentos estão detalhadamente descritos. Quanto mais detalhes sobre o processo de geração de dados, mais viável o julgamento sobre sua validade e qualidade, e a adoção de resultados obtidos por outros profissionais para usar em suas próprias pesquisas depende do resultado dessa avaliação. O conceito de autenticidade está vinculado à origem dos dados, ao profissional ou à instituição responsáveis pela coleta, critérios importantes, embora aparentemente menos relevantes. Enquanto os arquivistas se preocupam com o documento, os cientistas dão importância ao conteúdo. Os metadados, enquanto informações auxiliares, são essenciais para facilitar a avaliação que o cientista fará dos dados à sua disposição (ROEDER; EPPARD; UNDERWOOD; LAURIENT, 2008; WALLIS et al., 2007).

<sup>74</sup> No caso de o documento sofrer alterações, é imprescindível controlar quem alterou, quando e o que foi alterado.

<sup>75</sup> MacNeil e Gilliland-Swetland (2005a) defendem uma relação mais estreita entre autenticidade e integridade, uma relação segundo a qual a autenticidade de um documento está vinculada à sua identidade e integridade, sendo a identidade relacionada a atributos que distinguem um documento de outros: responsáveis pela criação; datas; indicação da ação ou processo da qual faz parte; relação com outros documentos; e anexos. Neste trabalho, o conceito de identidade é sinônimo de autenticidade, mantendo-se o conceito de integridade à parte.

<sup>76</sup> Transferência de ambiente tecnológico pode referir-se a uma simples transmissão via correio eletrônico ou ao procedimento de migração.

computadores diferentes. Nesse caso, temos apresentações diferentes da mesma sequência de bits, mas que necessariamente não são o “mesmo” documento (LEVY, 2000; SMITH, 2000b). As diferentes visualizações comprometem a integridade dos documentos digitais? Imagens digitais (fotografias ou a reprodução de um quadro), por exemplo, conforme a calibração do monitor utilizado para visualizá-las, podem sofrer variação no tom das cores, que muitas vezes são os elementos fundamentais de suas composições e cuja alteração pode interferir na percepção do observador. Um texto ou uma apresentação em *slides*, por exemplo, cuja formatação original se perdeu, pode ter a compreensão do seu conteúdo comprometida.

É fundamental estabelecer critérios que permitam avaliar a autenticidade de um documento e verificar se com o passar do tempo ele continua autêntico e íntegro. O criador e o guardião devem contar com ferramentas que sustentem as declarações sobre como, quando, por quê e por quem o documento foi criado; alterações que sofreu, feitas por quem, quando e por que; aspectos técnicos da digitalização ou criação; e outras informações relevantes. O criador deve fornecer declarações que depois de comprovadas são endossadas e reunidas com as declarações feitas pelo guardião.

A autenticidade e a integridade de documentos manuscritos são, aparentemente, mais fáceis de ser comprovadas do que as de documentos digitais, pois falta para estes evidências físicas, tão confiáveis ou estáveis quanto as de documentos analógicos, que auxiliem no processo de autenticação (CULLEN, 2000). O trabalho com documentos natodigitais exige muito mais cautela do que com documentos digitalizados, pois o original analógico pode ser considerado o ponto de partida para a verificação e constatação da autenticidade e da integridade do digital, desde que, obviamente, seja autêntico e íntegro por sua vez; lembrando mais uma vez que não há preocupação em verificar a veracidade do conteúdo dos documentos.

## **4.2 Estratégias para verificar a autenticidade e a integridade de documentos digitais**

Pelo ineditismo do trabalho com documentos digitais, pelo menos nas primeiras tentativas de estabelecer critérios que auxiliem na avaliação e comprovação de sua autenticidade e integridade, é interessante tentar aproveitar, com as devidas adaptações, o

conhecimento adquirido e a experiência desenvolvida com o trabalho com documentos analógicos. Um exemplo de metodologia tradicional é testar a autenticidade de um documento fazendo perguntas quanto à origem, ao autor, à data de criação, a um possível terceiro<sup>77</sup> a quem o documento já pertenceu e a outros atributos que sejam comprováveis e que permitam atestar sua autenticidade (CULLEN, 2000). Informações sobre a origem, de maneira geral, dizem respeito à história do documento, à cadeia de custódia e também à sua relação com outros documentos (LYNCH, 2000).

As técnicas propostas para a comprovação da autenticidade são similares às de verificação da integridade.

Para garantir a autenticidade dos documentos, as instituições responsáveis devem criar e documentar políticas e procedimentos para controlar a criação, o recebimento, a transmissão, a manutenção e a disposição dos documentos, protegendo-os de alterações, uso e, até mesmo, eliminação ilegais. [...] Para garantir a integridade é necessário que um documento esteja protegido contra alterações ilegais. Políticas de gerenciamento de documentos devem especificar quais alterações ou anotações, em quais circunstâncias e autorizadas por quem, podem ser feitas a um documento. Qualquer [ação] autorizada deve ser registrada e recuperável (ISO, 2001, p. 7).

Os documentos digitais, sejam os digitalizados, sejam, em especial, os natodigitais, dependem de informações fornecidas e preservadas por alguém<sup>78</sup>. É importante preservar essas informações sobre os documentos; informações sobre possíveis transformações e alterações sofridas, o histórico que possa ser rastreado e que permita a identificação da origem do documento ou, pelo menos, da versão que inspire autenticidade<sup>79</sup> (LEVY, 2000). Esses registros permitiriam avaliar se as transformações violaram de alguma forma a autenticidade ou a integridade do documento.

A canonicalização<sup>80</sup> é uma estratégia que propõe verificar a integridade de um documento submetido à migração de formato, por exemplo, comparando-o após o procedimento a um arquétipo (*canonical form*) estabelecido para o seu tipo documental (LYNCH, 1999). Ainda no sentido de evitar dúvidas quanto a alterações, sugere-se que, quando for necessária ou desejada uma correção ou atualização no conteúdo do documento, poderia ser adotada a errata, prática comum para documentos impressos, ou que poderia ser

---

<sup>77</sup> Terceiro é entendido como uma instituição que já tenha tido o documento em questão sob sua custódia.

<sup>78</sup> Essas informações são os metadados.

<sup>79</sup> A existência de várias versões de um mesmo documento demanda cautela. É necessário diferenciá-las, controlá-las e preservá-las satisfatoriamente. Podem também existir cópias de um mesmo documento de diferentes repositórios.

<sup>80</sup> Tradução literal do inglês *canonicalization*.

feito algum tipo de comunicação e/ou correção externa ao documento para não comprometer a confiança na integridade já atestada (SMITH, 2000b). Em alguns casos, a errata pode não ser viável, dependendo do erro identificado, então, um boa solução seria fazer a correção necessária e registrá-la no histórico do documento (quem, quando e por que o documento foi manipulado), conforme recomendações da ISO.

Para a verificação das informações sobre os documentos ou sua autenticidade e integridade, Andy Hopper, da Universidade de Cambridge (CULLEN, 2000), sugere que se confie nas declarações oferecidas por uma instituição ou profissional respeitáveis<sup>81</sup> (*trusted third party*), até que por algum motivo se prove que alguém se enganou em relação ao documento, invalidando assim o que se pensava saber sobre ele. Muitas pessoas preferem ou têm de confiar no veredicto de um especialista (TANSELLE, 2001) – é possível até deduzir que, mesmo que políticas e algumas informações técnicas sobre autenticidade e integridade de documentos e os procedimentos para sua verificação e preservação estejam à disposição dos usuários, muitos se contentarão com uma simples declaração oferecida pela instituição responsável pela custódia. Ou seja, talvez poucos usuários sintam necessidade de analisar as informações técnicas para então confiar no atestado de autenticidade e/ou<sup>82</sup> integridade que lhes for apresentado: a própria instituição é considerada o *selo de garantia*.

De qualquer forma, é importante que as políticas adotadas pelo guardião para verificar e preservar a autenticidade e a integridade dos documentos, sob sua custódia estejam disponíveis de maneira fácil e objetiva a todos. “Os usuários têm um importante papel na avaliação da natureza e do grau de confiança desses materiais em determinadas circunstâncias; [...] os procedimentos que bibliotecários e arquivistas estabelecem para a preservação da autenticidade de fontes digitais são apenas o começo [...]” (MACNEIL; MAK, 2007, p. 47).

Ainda não há consenso sobre o significado dos conceitos autenticidade e integridade, mas todos os profissionais envolvidos com bibliotecas, arquivos e informática têm consciência da importância da garantia desses atributos, principalmente por causa de importantes mudanças introduzidas pelas novas tecnologias. As técnicas para verificação também não estão estabelecidas, mas provavelmente serão baseadas em técnicas aplicadas aos documentos analógicos e na experiência dos profissionais da Arquivística.

Aqui foram apresentadas algumas estratégias, e no próximo tópico estão alguns

---

<sup>81</sup> Acreditamos que o gerenciamento cuidadoso e transparente de um acervo confere confiabilidade a uma instituição, além da tradição e do respeito conquistados perante sua comunidade.

<sup>82</sup> “E/ou” foi utilizado em razão da não coexistência necessária dos dois atributos em um mesmo documento, como foi comentado anteriormente.

elementos importantes a serem verificados na avaliação da autenticidade e integridade de documentos digitais.

### **4.3 Elementos para verificar a autenticidade e a integridade de documentos digitais**

Muitas iniciativas se baseiam na experiência secular em lidar com documentos analógicos e no trabalho com documentos digitais em arquivos, mesmo assim, mediante as adaptações necessárias, suas propostas podem ser aplicadas a repositórios e patrimônios de qualquer área. No âmbito acadêmico, o patrimônio a ser preservado compreende toda a produção intelectual da universidade e também a de terceiros, que pode ser encontrada no acervo de suas bibliotecas físicas.

O grupo de trabalho sobre autenticidade do Projeto InterPARES<sup>83</sup> identificou elementos essenciais para verificação e manutenção da autenticidade de documentos digitais, elementos estabelecidos com base em considerações sobre quais deles seriam comuns a todos os documentos, quais poderiam diferenciar os documentos entre si e quais permitiriam verificar a autenticidade em seu tempo e posteriormente. Esses elementos<sup>84</sup> foram divididos em duas categorias:

- **série A**, para verificação da autenticidade antes de o documento ser transferido do seu criador para o guardião<sup>85</sup>.

---

<sup>83</sup> InterPARES significa Investigação Internacional sobre Documentos Autênticos Permanentes em Sistemas Eletrônicos. Foi criado em 1999 com base no crescente interesse nos resultados de outro projeto: *Preservation of the integrity of electronic records*, desenvolvido pela *School of Library, Archival and Information Studies*, da *University of British Columbia* (DURANTI, 2005). “O projeto tem por objetivo desenvolver conhecimentos essenciais para a preservação de documentos digitais autênticos e fornecer padrões, políticas, estratégias e planos de ação capazes de garantir a longevidade desse material e a possibilidade de os usuários confiarem na sua autenticidade” (InterPARES, *on-line*).

<sup>84</sup> Para identificar esses elementos, o grupo de trabalho se baseou em uma análise e adaptação contemporânea da diplomática arquivística tradicional, em que “há quem defenda que a preservação do material no seu estado original deverá ser considerada a única medida de sucesso” para garantir que ele não sofreu alterações (FERREIRA, 2006, p. 64).

<sup>85</sup> Em alguns casos não existem informações suficientes para que todos os requisitos da série A sejam atendidos. Nesses casos, o guardião deve utilizar todas as informações disponíveis e conhecidas sobre o processo de criação, gerenciamento e manutenção, para avaliar a autenticidade do documento que lhe foi entregue. Pode ainda valer-se de outras cópias, cópias de segurança, trilhas de auditoria e declaração de um terceiro de confiança (MACNEIL; GILLILAND-SWETLAND, 2005a).

- **série B**, para verificação da manutenção da autenticidade<sup>86</sup> no decorrer do tempo.

Os elementos da **série A** foram extraídos da noção de um sistema confiável para preservação de documentos<sup>87</sup>, um sistema que se apóia em regras que norteiem a seleção, o controle de acesso, as ações com os documentos e o gerenciamento do ciclo de vida deles até sua destinação final, caso não tenha sido determinada sua preservação permanente (MACNEIL; GILLILAND-SWETLAND, 2005a). A série A é subdividida em:

- Elementos sobre o contexto imediato da criação do documento (data e responsáveis); como tem sido gerenciado; sua identidade; e fundamentos para demonstrar sua integridade; e
- Elementos sobre controles de procedimento de criação, tratamento e manutenção dos documentos que permitem inferir sua integridade (MACNEIL; GILLILAND-SWETLAND, 2005b).

Os elementos da **série B** foram extraídos da noção de guardião confiável que deve demonstrar sua capacidade de implantar e seguir procedimentos de segurança e preservação da integridade dos documentos sob sua custódia; e demonstrar que não tem interesse em alterar nem permitir que ninguém altere os documentos (MACNEIL; GILLILAND-SWETLAND, 2005a). A série B é subdividida em:

- Elementos sobre políticas de gerenciamento;
- Elementos sobre os processos de reprodução; e
- Elementos sobre o contexto de criação, procedência e alterações sofridas desde a criação do documento (MACNEIL; GILLILAND-SWETLAND, 2005b).

Para continuar a reflexão, é importante esclarecer que em alguns textos há certa variação na escolha do termo empregado para se referir aos elementos apresentados. Em geral, o termo adotado é **metadados**<sup>88</sup>, informações úteis que exigem controle e preservação para que possam ser empregadas na tarefa de verificar a autenticidade e a integridade de documentos digitais no decorrer do tempo. Quanto maior a quantidade de boas informações

---

<sup>86</sup> Acredito que, uma vez provada a autenticidade de um documento, não haverá mudanças a menos que se identifique alguma falha no processo de avaliação, ao passo que a integridade, sim, precisa ser avaliada constantemente. No entanto, como estou apresentando propostas de um grupo de pesquisa, devo respeitar a terminologia adotada por ele.

<sup>87</sup> O próximo capítulo é dedicado à questão dos repositórios digitais confiáveis.

<sup>88</sup> Definição apresentada no tópico 3.2.



sobre um documento, mais forte será a presunção de autenticidade (MACNEIL; GILLILAND-SWETLAND, 2005a) e integridade.

Com base nas séries de elementos apresentadas anteriormente (séries A e B), proponho alguns metadados para o controle de documentos acadêmicos natodigitais<sup>89</sup>:

- Autores
- Data de criação
- Local de criação
- Gênero documental (audiovisual, textual, iconográfico, cartográfico, multimídia)<sup>90</sup>
- Data de depósito no repositório digital
- Informações descritivas (título, assunto, descrição física)
- Finalidade (relatório, evento, aula e outros)
- Restrições de uso e controle de direitos autorais
- Formato do arquivo digital e respectiva versão
- Resolução da imagem
- Tamanho do arquivo digital
- Responsável pela verificação das informações (profissional autorizado)
- Data da última verificação
- *Software* utilizado (informações sobre a versão e o tipo)
- Estratégia de preservação
- Data do último procedimento para preservação
- Responsável pelo procedimento de preservação (quando não for uma ação realizada automaticamente pelo sistema)
- Anotações sobre possíveis problemas identificados no documento (alteração de fonte, corrupção ou perda de dados, por exemplo)
- Medidas tomadas a partir da identificação do problema (correções)
- Responsável pelas medidas tomadas com base na identificação do problema
- Data das correções

Para documentos digitalizados poderiam ainda ser acrescentadas informações sobre:

- Técnica de digitalização

---

<sup>89</sup> Para sugerir esses metadados, não foram considerados apenas o acervo das bibliotecas digitais de teses e dissertações, mas uma produção acadêmica ampla.

<sup>90</sup> Indica “a configuração da informação no documento de acordo com o sistema de signos utilizado” (CONSELHO NACIONAL DE ARQUIVOS, 2009, p. 139).

- Equipamentos utilizados para digitalização
- Data da digitalização
- Responsável pela digitalização
- Responsável por comparar o produto final com o documento analógico
- Declaração sobre a autenticidade e integridade do documento analógico original

O objetivo desta seção não é esgotar as possibilidades de metadados, mas os próximos a ser apresentados foram extraídos do Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos (e-ARQ Brasil)<sup>91</sup> (CONSELHO NACIONAL DE ARQUIVOS, 2009) para detalhar um pouco a lista anterior:

- Identificador do documento (deve ser único e persistente, gerado automaticamente)
- Identificador do componente digital (relacionamento com seus componentes, por exemplo, no caso de documentos multimídia compostos de arquivos de áudio e vídeo)
- Indicação de anexos
- Relação com outros documentos
- Características técnicas do componente digital (nível de composição – compressão ou criptografia –, tamanho, software de criação e inibidores – restrição de acesso, uso ou migração)
- Armazenamento (localização e suporte de armazenamento)
- *Hardware* (modelo, versão, fabricante, tipo – por exemplo, memória, processador –, configuração mínima para instalar e operar e documentação)
- Fixidade (para verificar alterações não documentadas ou autorizadas)
- Validação (identificação da validação, agente responsável e detalhes)

Seria interessante um trabalho cooperativo entre bibliotecários ou responsáveis pelo repositório digital, caso esteja vinculado a outro setor da universidade, e autores para garantir que os documentos sejam criados conforme os padrões. O padrão de composição de documentos garante a existência de informações essenciais e pode facilitar a captura de metadados conforme recomendação do e-ARQ Brasil. O usuário precisa ter consciência da

---

<sup>91</sup> Esse modelo será apresentado no próximo capítulo.

utilidade e importância de cada metadado, e para isso é fundamental uma atuação transparente do guardião.

## 5 REPOSITÓRIOS DIGITAIS CONFIÁVEIS PARA PRESERVAÇÃO DA AUTENTICIDADE E DA INTEGRIDADE DE DOCUMENTOS DIGITAIS

As universidades são um dos principais centros de produção de conhecimento no Brasil, contudo parece que elas ainda não se preocupam com a preservação sistemática de seu patrimônio científico (MARTINS, R., 2006). As universidades e outras instituições são fundamentais para a preservação do patrimônio científico e tecnológico brasileiro composto de

acervos de documentos escritos originados de instituições científicas e de ensino, coleções organizadas por estudiosos, registros fonográficos e fotográficos, filmes, obras raras, máquinas e equipamentos, edifícios e instalações, bibliotecas, mapotecas, arquivos pessoais de pesquisadores e professores, parcela substancial da correspondência diplomática, documentos e instrumentos relativos a numerosas atividades militares e registros da produção científica (CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO, 2003, p.6)<sup>92</sup>.

Documentos científicos devem ser entendidos como qualquer forma de registro sobre um processo de pesquisa e os resultados obtidos, seja um registro formal (um artigo científico publicado oficialmente, por exemplo) ou informal (e-mail ou fóruns de discussão, por exemplo), independentemente da sua forma de apresentação: dados, gráficos, textos ou representações específicas (estruturas químicas, por exemplo) (EUROPEAN TASK FORCE PERMANENT ACCESS, 2005).

Nos últimos anos surgiram muitos projetos para criação de bancos de informação para divulgar à comunidade acadêmica e ao público em geral o rico acervo das universidades, centros de pesquisa e instituições culturais. Porém, não é suficiente disponibilizar acesso remoto via internet; o repositório digital precisa demonstrar a seus usuários que ele é confiável e capaz de preservar o patrimônio digital pelo qual é responsável, e para tanto deve

---

<sup>92</sup> Na seção de documentos adicionais sobre a proposta para o desenvolvimento da área de história da ciência no Brasil, no site do Grupo de História e Teoria da Ciência da Unicamp, declara-se sobre o relatório da comissão responsável: “Esse documento não está ainda disponível nos ‘sites’ do CNPq ou do MCT. Está disponibilizada aqui uma versão parcial, que foi divulgada no ano passado entre os pesquisadores” (Disponível em: <<http://ghtc.ifi.unicamp.br/SBHC/Proposta-SBHC-docs.htm>>. Acesso em: 15 maio 2010). No *site* do CNPq e do Ministério da Ciência e Tecnologia (MCT) não foi encontrada referência à política nacional de memória da ciência e da tecnologia, apenas à norma que criou a comissão responsável.

começar por expor o seu comprometimento com a missão de garantir acesso e preservação a longo prazo de documentos digitais igualmente confiáveis (RESEARCH LIBRARIES GROUP, 2002).

Determinar a confiança em um repositório exige mais do que avaliar a organização para preservação, inclui avaliar a instituição como um todo, sua organização, equipe, políticas e procedimentos, sustentabilidade financeira, licenças e responsabilidades sob as quais deve trabalhar (CENTER FOR RESEARCH LIBRARIES, 2007; ONLINE COMPUTER LIBRARY CENTER). O processo de avaliação e certificação da confiabilidade de um repositório é detalhado e complexo, são muitos os requisitos que devem ser atendidos pelos que almejam o *status* de confiável, mas podemos destacar alguns (CENTER FOR RESEARCH LIBRARIES, 2007; RESEARCH LIBRARIES GROUP, 2002):

- Aceitar a responsabilidade da preservação a longo prazo em nome dos depositantes e em consideração a usuários atuais e futuros;
- Criar sistemas conforme padrões amplamente aceitos para garantir a continuidade do gerenciamento, do acesso e da segurança do acervo;
- Estabelecer metodologias de avaliação do sistema que coincidam com as expectativas de confiabilidade dos usuários;
- Atuar de forma transparente;
- Garantir a integridade, autenticidade e usabilidade dos documentos;
- Criar e manter metadados para controle de ações de preservação, assim como nos contextos de produção, acesso e uso;
- Ter infraestrutura tecnológica adequada para garantir manutenção e segurança contínuas;
- Ter políticas, práticas e desempenho que possam ser mensurados e avaliados.

## **5.1 Parâmetros para garantir a confiabilidade de repositórios digitais**

Como ocorre em qualquer avaliação, para determinar o grau de confiabilidade de um repositório são necessários parâmetros que sirvam também para orientar sua construção e gerenciamento. Existem iniciativas nacionais e internacionais dedicadas a estabelecer esses parâmetros, porém ainda não há consenso. No Brasil, aparentemente, não existem organismos

credenciados para fazer auditoria em Sistemas de Gestão em Tecnologia da Informação e conceder certificados de qualidade<sup>93</sup>. É aceitável a auto avaliação de uma instituição seguindo padrões e critérios amplamente aceitos, porém o ideal é combiná-la com a auditoria realizada por terceiros (RESEARCH LIBRARIES GROUP, 2002).

Buscando parâmetros básicos para repositório e os específicos para preservação e controle da autenticidade e integridades dos documentos, primeiramente apresentamos alguns dos requisitos propostos pelo e-ARQ Brasil<sup>94</sup>. Optou-se por manter a numeração original de cada item para que os interessados possam localizá-los facilmente no documento-fonte:

### **1 Organização dos documentos arquivísticos: plano de classificação e manutenção dos documentos**

**1.1.36** Impedir a eliminação de um documento ou de qualquer parte de seu conteúdo, a menos que determinado em algum instrumento de auxílio ao gerenciamento (por exemplo, tabela de temporalidade e destinação de documentos).

**1.1.65, 6.4.1 e 8.3.1** (Requisitos de Segurança e Preservação, respectivamente) Registrar na trilha de auditoria informações sobre data e hora da captura; responsável pela captura; data e hora de produção, acréscimo e eliminação de metadados; alterações efetuadas nas permissões de acesso; tentativas de exportação e importação (incluindo *backup* e *restore*); usuário, data e hora de acesso a documentos; eliminação do documento e seus metadados; infrações cometidas contra mecanismos de controle de acesso; operações de preservação realizadas; ações administrativas relativas a usuários (cadastro, ativação, bloqueio, atualização de dados e permissões, troca de senha etc.); eventos de administração e manutenção das trilhas de auditoria (alarmes, cópias, configuração de parâmetros etc.); entre outras.

---

<sup>93</sup> O Instituto Nacional de Metrologia, Normalização Qualidade Industrial (Inmetro) é responsável por credenciar Organismos de Avaliação da Conformidade. A busca realizada por *Tipo de organismo* no banco de informações – <http://www.inmetro.gov.br/organismos/index.asp> – não recuperou nenhum para Sistemas de Gestão em Tecnologia da Informação. O Inmetro é membro do *International Accreditation Forum* (IAF) e recebeu em 1999 o *status* de confiança entre os países signatários do Memorando de Entendimento, cujo objetivo é facilitar o comércio e a aceitação de certificados de qualidade de produtos e serviços emitidos pelos países-membros (Disponível em: <<http://www.inmetro.gov.br/barreirastecnicas/organismosRelacionados.asp>>. Acesso em: 31 maio 2010).

<sup>94</sup> O e-ARQ foi elaborado pela Câmara Técnica de Documentos Eletrônicos do Conselho Nacional de Arquivos e, se for utilizado no desenvolvimento e gerenciamento de um repositório, “conferirá credibilidade à produção e à manutenção de documentos” digitais, garantindo também sua confiabilidade, autenticidade e acessibilidade (CONSELHO NACIONAL DE ARQUIVOS, 2009, p. 9). Na escolha dos requisitos, não foram considerados os que são específicos para gestão de arquivos, mas apenas alguns dos que foram julgados aplicáveis a qualquer repositório digital.

## **2 Tramitação e fluxo de trabalho**

**2.1.12 e 2.1.20** Fornecer histórico de movimentação dos documentos em metadados, por exemplo, referentes à data e hora de entrada e saída, e nomes de responsáveis.

## **3 Captura**

**3.1.12, 3.1.13 e 6.10.4** (Requisito de Segurança) Garantir que os metadados de um documento sejam inseridos e alterados somente por administradores e usuários autorizados e as ações devidamente registradas em trilhas de auditoria.

**3.1.15** Inserir automaticamente o maior número possível de metadados previstos no sistema para diminuir as tarefas do usuário e garantir maior rigor na inserção.

## **6 Segurança**

**6.1.1 e 6.1.5** Criar cópias de segurança periódicas de todos os documentos, metadados e parâmetros do sistema conforme agendamento do administrador.

**6.1.9** Copiar com frequência as trilhas de auditoria e armazenar em pelo menos um local diferente de onde fica o repositório (*off-site*).

**6.2.5** Permitir acesso a funções do sistema somente a usuários autorizados e sob controle rigoroso da administração para proteger a autenticidade dos documentos.

**6.4.3 e 6.4.6** Impedir qualquer modificação na trilha de auditoria e permitir apenas ao administrador e ao auditor a sua leitura.

**6.4.4** Assegurar que as informações da trilha de auditoria estejam disponíveis para inspeção, permitindo que uma ocorrência específica possa ser identificada, e assegurar que todas as informações sejam compreensíveis.

## **7 Armazenamento**

**7.1.1** Utilizar, preferencialmente, dispositivos e padrões de armazenamento maduros, estáveis no mercado e amplamente disponíveis.

**7.1.3** Efetivar atividades de migração sempre que se torne evidente ou previsível a obsolescência do padrão corrente.

**7.2.1 e 7.2.2** Possuir capacidade de armazenamento suficiente para acomodação dos documentos e suas cópias de segurança e prever a possibilidade de expansão da estrutura de armazenamento.

**7.3.3** Utilizar mecanismos de proteção que previnam alterações indevidas e mantenham a integridade dos dados armazenados.

**7.3.5** Verificar periodicamente a integridade.

## 8 Preservação

**8.1.3 e 8.1.4** Controlar a vida útil dos suportes para auxiliar no processo de atualização e contar com funcionalidades que informem quais são os suportes cuja vida útil se encontra perto do fim.

**8.2.2 e 8.2.6** Verificar periodicamente os dados armazenados para detectar possíveis erros (recomenda-se o uso de códigos de verificação – *checksums*) e armazenar o histórico dos resultados de verificação.

**8.3.4 e 8.3.5** Adotar normas amplamente aceitas, descritas em especificações abertas e disponíveis publicamente em vez de estruturas proprietárias referente a estruturas para codificação, armazenamento e banco de dados.

**8.3.6** Gerenciar metadados relativos à preservação dos documentos e seus respectivos componentes.

Há semelhanças entre as propostas de cada iniciativa. Para escolher alguns dos requisitos propostos pela iniciativa internacional<sup>95</sup>, a preferência foi pelos que complementam os nacionais apresentados e por alguns semelhantes visando a demonstrar a importância dos aspectos abordados e a preocupação mundial a respeito deles. Para os requisitos internacionais também foi mantida a numeração original de cada um para que os interessados possam localizá-los facilmente no documento-fonte:

### A Infraestrutura organizacional

**A1.1** Apresentar de forma detalhada, clara e acessível a depositantes e demais interessados a missão e uma declaração de comprometimento com a manutenção a longo prazo.

**A2.3** Capacitar a equipe profissional para enfrentar os desafios impostos pela evolução tecnológica.

**A3.2** Registrar políticas e procedimentos, e disponibilizar em formato tangível. Prever revisão e atualização conforme mudanças tecnológicas, crescimento do repositório ou mudança nas práticas de trabalho.

**A3.7** Comprometer-se com a transparência e explicação de todas as ações relacionadas ao gerenciamento, em especial as relacionadas à preservação digital.

**A3.8** Desenvolver e adaptar medidas apropriadas para garantir a integridade dos documentos sob sua custódia.

---

<sup>95</sup> Critérios extraídos do relatório elaborado pelo *Online Computer Library Center* e *Center for Research Libraries* (2007). Na lista do relatório constam mais de 80 critérios.



## **B Gerenciamento de objetos**

**B1.3** Garantir a autenticidade dos documentos.

**B1.2** e **B1.4** Especificar as informações que devem acompanhar o documento desde o depósito e verificar cada objeto para garantir a completude e correção em comparação às especificações.

**B3.1** Registrar as estratégias de preservação.

**B4.4** Monitorar a integridade por meio de mecanismo de autenticação. Arquivar separadamente os códigos de verificação, informações que os relacionam a um pacote de metadados e o respectivo pacote. Documentar como os *checksums* e metadados são armazenados separadamente<sup>96</sup>.

## **C Tecnologias, infraestrutura técnica e segurança**

**C1.2** Garantir *hardware*, *software* e processos adequados para *backup*.

**C1.5** e **C1.6** Detectar com eficiência a corrupção ou perda de dados de qualquer origem. Relatar e reparar com eficiência as violações de integridade. Documentar os procedimentos para reparação e avaliação do sucesso do procedimento.

**C1.7** Definir prazos e processos para mudança de mídia de armazenamento e/ou *hardware*.

**C2.1** e **C2.2** Garantir *hardware* e *software* apropriados para dar apoio ao serviço prestado e garantir procedimentos adequados para monitorar notificações e avaliar a necessidade de mudanças tecnológicas.

**C3.3** Atribuir funções, responsabilidades e autorizações para os membros da equipe<sup>97</sup>.

**C3.4** Documentar um plano adequado de prevenção e recuperação contra desastres, incluindo, pelo menos, um *backup* geral e uma cópia do plano armazenados em local diferente da sede do repositório.

O responsável pela preservação de um patrimônio será respeitado se demonstrar ter controle e organização. A transparência é a melhor maneira de provar que o repositório é gerenciado de acordo com padrões e práticas consensuais e que o guardião está trabalhando para cumprir sua missão.

---

<sup>96</sup> Recomendações do grupo de pesquisa sobre metadados de preservação baseadas no modelo do *Open Archival Information System* (OAIS) (ONLINE COMPUTER LIBRARY CENTER, 2002).

<sup>97</sup> As autorizações determinam quem pode fazer o que no sistema, incluir usuário, alterar metadados, ter acesso aos códigos de verificação (*checksums*) (CENTER FOR RESEARCH LIBRARIES, 2007; ONLINE COMPUTER LIBRARY CENTER).

## 6 BIBLIOTECAS DIGITAIS DE TESES E DISSERTAÇÕES DAS UNIVERSIDADES ESTADUAIS PAULISTAS

Há alguns anos, a Universidade de São Paulo (USP), a Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp) e a Universidade Estadual de Campinas (Unicamp) buscam garantir às suas comunidades acesso às principais fontes de informação científica nacionais e internacionais: acesso físico ou remoto a um “universo [...] mais restrito e controlado [que compreende] catálogos das bibliotecas do sistema, bases de dados, revistas eletrônicas e livros eletrônicos (*e-books*), além de *links* para portais de interesse acadêmico” (MARQUES, 2008-2009, p. 22).

Quanto à produção intelectual e ao patrimônio documental, as três universidades têm desenvolvido projetos para divulgar, para sua comunidade interna e para a sociedade em geral, os conhecimentos produzidos com a finalidade de “aumentar a visibilidade e facilitar o aproveitamento desse conhecimento nas atividades realizadas fora do meio estritamente acadêmico” (KONDO; LIRANI; TRAINA JR., 2008-2009, p. 70) – esses projetos são instrumentos de apoio cultural e científico.

Dentre os vários projetos universitários para a criação de repositórios, foram escolhidas as bibliotecas digitais de teses e dissertações porque são importantes resultados da produção intelectual dessas instituições e porque desde 2006 a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), subordinada ao Ministério da Educação, tornou obrigatória a divulgação desses trabalhos na internet<sup>98</sup>. Neste estudo, a escolha pelas bibliotecas digitais da USP, Unesp e Unicamp se deve à parceria entre essas instituições e a alguns aspectos pioneiros de seus projetos.

A **Biblioteca Digital de Teses e Dissertações da USP** foi inaugurada em 2001 adotando padrões internacionais; em 2002, começou a integrar a Biblioteca Digital Brasileira de Teses e Dissertações; e, em 2003, filiou-se à NDLTD<sup>99</sup>. Essa biblioteca foi modelo para o desenvolvimento de projetos similares em outras instituições em razão de seu caráter pioneiro

---

<sup>98</sup> BRASIL. Ministério da Educação. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. Portaria nº 13, de 15 de fevereiro de 2006, que institui a divulgação digital das teses e dissertações produzidas pelos programas de doutorado e mestrado reconhecidos. Disponível em: <[http://www.capes.gov.br/images/stories/download/legislacao/Portaria\\_013\\_2006.pdf](http://www.capes.gov.br/images/stories/download/legislacao/Portaria_013_2006.pdf)>. Acesso em: 10 fev. de 2010.

<sup>99</sup> *Networked Digital Library of Theses and Dissertations*, já apresentada anteriormente.

e da dimensão do sistema de pós-graduação da universidade (KONDO; LIRANI; TRAINA JR., 2008-2009). Seu acervo, até 17 de maio de 2010, contava com 22.890 trabalhos<sup>100</sup>.

A **Biblioteca Digital de Teses e Dissertações da Unesp** começou a funcionar em 2003 utilizando o *software* livre Nou-Rau, desenvolvido pela Unicamp. Também está filiada à NDLTD (FOGOLIN; KEMPINAS, [2004]) e recentemente se incorporou à Biblioteca Digital Brasileira de Teses e Dissertações. Seu acervo, até 17 de maio de 2010, contava com um total de 9.609 trabalhos<sup>101</sup>.

A **Biblioteca Digital de Teses da Unicamp** foi criada em 2001 e hoje é parte da Biblioteca Digital da Universidade<sup>102</sup>. De 2004 a 2009, foi realizado um grande projeto de digitalização retrospectiva e por isso a Unicamp é a primeira das três universidades a oferecer toda a sua produção de teses e dissertações em meio eletrônico e com acesso controlado por meio do cadastro obrigatório dos usuários antes de visualizar ou fazer o *download* do trabalho de interesse (SUGIMOTO, 2009). Seu acervo, até 17 de maio de 2010, contava com 32.245 trabalhos.

Neste estudo exploratório, conhecer as práticas das bibliotecas digitais para avaliar e atestar a autenticidade e a integridade dos documentos em seu acervo permitiria constatar se as recomendações apresentadas na literatura são adotadas e identificar a necessidade de melhorias. Para a coleta dos dados, optou-se pela aplicação de um questionário via e-mail ao diretor do Sistema de Bibliotecas da Unicamp, à coordenadora geral de Bibliotecas da Unesp e à diretora técnica do Sistema Integrado de Bibliotecas da USP, que recomendou o encaminhamento do questionário ao Centro de Informática de São Carlos (CISC/USP). No tópico a seguir, as respostas de cada instituição sobre as práticas de preservação do acervo de suas bibliotecas digitais são apresentadas na ordem das perguntas do questionário, agrupadas por tema para facilitar a comparação entre elas. As respostas, tal como encaminhadas, encontram-se anexas, com a concordância dos respondentes.

---

<sup>100</sup> Do total, 11.219 (de 2007 a 2009, pois na data da pesquisa ainda não haviam sido incorporados trabalhos defendidos em 2010) são trabalhos concluídos a partir de 2007, ano em que passou a ser obrigatório o depósito na biblioteca digital (dados coletados em 20 de julho de 2009). Os trabalhos concluídos anteriormente são depositados espontaneamente pelo autor, que assina um termo de autorização. Em ambas as situações, a versão digital deve ser idêntica à impressa que está no acervo da biblioteca da unidade correspondente.

<sup>101</sup> Cada biblioteca (USP e Unicamp) apresenta suas estatísticas diferentes. A Unesp, por sua vez, não apresenta nenhum dado, o total do acervo foi determinado a partir do resultado de busca feita no catálogo. Nesse caso, para tentar recuperar a maior quantidade possível de trabalhos realizados, utilizou-se busca pelo tipo de trabalho no campo Notas, fixando o período de publicação entre 1900 e 2010.

<sup>102</sup> Alternativa que a comunidade acadêmica tem para disponibilizar a sua produção na internet, além de teses, “artigos, fotografias, ilustrações, obras de arte, revistas, registros sonoros, teses, vídeos e outros documentos de interesse ao desenvolvimento científico, tecnológico e sócio-cultural” (SBU, *on-line*).

## 6.1 A prática das universidades estaduais paulistas: síntese das respostas ao questionário e algumas considerações

Nenhuma das três universidades tem uma política de preservação formalmente estabelecida, embora reconheçam sua importância. A Unicamp tem um grupo de estudos envolvido com o Arquivo Nacional e o Projeto InterPARES visando a estabelecer padrões para gestão, preservação e acesso a documentos arquivísticos que poderão ser adotados para os demais repositórios da universidade, incluindo a biblioteca digital de teses e dissertações. A Unesp também manifestou a intenção de desenvolver uma política, e a USP não se manifestou.

Mesmo sem ter uma política formal, todas as instituições adotaram estratégias de preservação conforme recomendações internacionais e práticas de instituições consideradas modelo. As estratégias mencionadas são *backup* e migração de formatos, *hardware* e *software* quando necessário. As universidades também adotaram padrões internacionais de metadados (MARC e Dublin Core) e o modelo OAIS na estrutura de suas bibliotecas digitais, além da parceria com iniciativas internacionais para aprimorar suas práticas.

Em relação aos conceitos de autenticidade e integridade, nenhuma das universidades possui definições formalmente estabelecidas. A Unesp foi a única que explicou o que entende por cada um dos conceitos com definições que se aproximam das que foram apresentadas neste trabalho, e reforçou a importância de garantir a integridade e contar com uma estrutura que permita identificar possíveis alterações sofridas por um documento após sua criação. Os estudos da Unicamp com o InterPARES estão em estágio avançado, mas a instituição ainda não adotou nenhuma definição, e a USP não se pronunciou a respeito.

Mesmo sem ter uma política formalizada, as universidades adotaram estratégias de preservação, e, da mesma forma, sem ter definições para os conceitos de autenticidade e integridade, adotaram estratégias para verificar esses atributos nos documentos que serão inseridos no acervo das bibliotecas digitais. As seções de pós-graduação e os bibliotecários são os responsáveis por verificar as informações fornecidas pelos autores no momento da submissão e por conferir a compatibilidade da versão digital com a versão impressa, verificando a integridade do documento. Não existem, no entanto, estratégias para verificar a preservação da integridade com o passar do tempo, pois, segundo as universidades, a restrição de acesso aos arquivos e metadados é suficiente para evitar alterações. A Unicamp relatou

que, quando é necessário fazer alguma alteração, a data em que isso ocorreu fica registrada nos metadados. Talvez a USP e a Unesp também tenham um procedimento semelhante para controle, mas não o explicitaram em suas respostas. Quando surgem dúvidas sobre a integridade de um arquivo eletrônico, elas são solucionadas por meio da comparação com o original impresso depositado na respectiva biblioteca.

Sobre a certificação da autenticidade e da integridade dos documentos no acervo das bibliotecas digitais, nenhuma das universidades explicita aos usuários suas práticas nem oferece atestados. A USP e a Unicamp consideram que a tramitação interna por setores, como a Comissão de Pós-graduação e a biblioteca, e a própria disponibilidade do trabalho na biblioteca digital são as garantias de que o documento digital é confiável. A Unesp apenas declarou não haver demonstração explícita.

Como não existem políticas formais, também não há uma periodicidade para a revisão dos procedimentos e das rotinas de trabalho, que são alteradas conforme a necessidade, e quando isso ocorre são registradas novas instruções de serviço, normas e resoluções. A USP e a Unesp declaram divulgar todas as normas<sup>103</sup>. A Unicamp afirma que não divulga sua política porque ela ainda está em fase de elaboração.

Não há a pretensão de oferecer soluções nem de criticar as iniciativas das universidades e de outras instituições que buscam preservar seu acervo e aprimorar suas práticas. Mesmo sem condições de realizar uma auditoria nas bibliotecas digitais, nos moldes do trabalho realizado pelo *Center for Research Libraries (CRL)*<sup>104</sup>, elaborei um quadro com os parâmetros para repositórios digitais apresentados no capítulo 5, reagrupados por tema, comparando-os com as respostas recebidas sobre as práticas correntes nas três universidades estaduais paulistas.

---

<sup>103</sup> Talvez a formulação da pergunta número 15 tenha ficado um pouco confusa, ou o que as universidades entendem por política não coincide com o que se encontra na literatura e na definição apresentada anteriormente na nota 34.

<sup>104</sup> O CRL concedeu ao projeto Portico, em janeiro deste ano, o certificado de repositório digital confiável, apesar de não cumprir perfeitamente todos os parâmetros recomendados; ressaltando que alguns deverão ser melhorados. O CRL acompanhará o desempenho para garantir a continuidade do bom serviço. O relatório público está disponível para consulta em:

<<http://www.crl.edu/sites/default/files/attachments/pages/CRL%20Report%20on%20Portico%20Audit%202010.pdf>>. Acesso em: 14 jun. 2010. Está em andamento a avaliação da iniciativa LOCKSS e de projetos chamados pelo CRL de esforços regionais. O Portico ([www.portico.org](http://www.portico.org)) é um serviço de preservação oferecido por ITHAKA, organização sem fins lucrativos dedicada a ajudar a comunidade acadêmica a usar as tecnologias digitais para preservar registros acadêmicos e avançar em pesquisas e ensino de maneira sustentável. Ele foi criado em 2002 como um projeto da *Andrew W. Mellon Foundation* para construir um arquivo digital sustentável a serviço da comunidade acadêmica e apoiar editoras e bibliotecas no começo de suas atividades com conteúdos digitais. O Portico é um repositório centralizado especializado na preservação de periódicos, *e-books* e outras produções. A USP, Unesp e Unicamp constam na lista de instituições brasileiras participantes.

Parâmetros para repositórios digitais confiáveis	USP	Unesp	Unicamp
<u>Controle de acesso</u>			
Utilizar mecanismos que previnam ações indevidas. Atribuir funções, responsabilidades e autorizações para os membros da equipe (critérios 1.1.36, 6.2.5, 7.3.3, C3.3)	Restringe acesso aos arquivos e metadados.	O tema não foi detalhado na resposta. Aparentemente existe controle de acesso aos arquivos.	Não foi mencionada a restrição de acesso, apenas o controle de ações dos usuários por meio do cadastro obrigatório (trabalhos visitados, <i>downloads</i> realizados, data e hora das ações, e IP da máquina utilizada para acesso).
<u>Trilha de auditoria</u>			
Registrar todas as informações sobre um documento. Impedir modificações (critérios 1.1.65, 6.4.1, 6.4.3, 6.4.6, 8.3.1, 2.1.12, 2.1.20, 6.4.4)	Não detalhou os metadados para gestão administrativa.	Não menciona.	Fica registrada a data em que foi realizada alguma alteração no arquivo depositado.
<u>Metadados</u>			
Especificar as informações que devem acompanhar o documento desde o momento do depósito e inserir automaticamente o maior número possível de metadados. Restringir acesso e manipulação dos metadados a usuários autorizados (critérios 3.1.12, 3.1.13, 6.10.4, 3.1.15, 8.3.6, B1.2, B1.4)	Os metadados descritivos são gerados por bibliotecários em formato Dublin Core. Os metadados administrativos são gerados pelo sistema.	O tema não foi detalhado na resposta. Adoção dos padrões MARC e Dublin Core.	Os metadados descritivos em formato MARC são originários do catálogo bibliográfico ACERVUS. Os metadados administrativos e estruturais são gerados pelo sistema.

Parâmetros para repositórios digitais confiáveis	USP	Unesp	Unicamp
<p><u>Estratégias de preservação</u></p> <p>Adotar normas amplamente aceitas e com especificações abertas. Efetivar atividades de migração e criar periodicamente cópias de segurança de todos os documentos, metadados, trilhas de auditoria e parâmetros do sistema, armazenando off-site. Capacitar a equipe (critérios 6.1.1, 6.1.5, 6.1.9, 7.1.1, 7.1.3, 8.3.4, 8.3.5, A2.3).</p>	<p><i>Backup</i> em disco e cartucho, espelhamento em equipamento de <i>storage off-site</i>, migração para <i>software</i> e <i>hardware</i> mais modernos.</p>	<p><i>Upgrade</i> de <i>softwares</i> e formatos, migração para plataformas mais modernas (<i>hardware</i>) e aquisição de <i>backfiles</i> e arquivos perpétuos de coleções digitais.</p>	<p><i>Backup</i> diário de todo o banco em diversos formatos, inclusive do <i>software</i> gerenciador. Gravação em duas cópias armazenadas em locais fisicamente separados.</p>
<p><u>Hardware e software</u></p> <p>Possuir capacidade de armazenamento suficiente para acomodar os documentos e suas cópias de segurança. Controlar a vida útil dos suportes e definir prazos e processos para atualizações tecnológicas (critérios 7.2.1, 7.2.2, 8.1.3, 8.1.4, C1.2, C1.7)</p>	<p>O controle da vida útil dos equipamentos foi pouco detalhado na resposta.</p>	<p>Há preocupação com atualização tecnológica.</p>	<p>Não menciona.</p>

Parâmetros para repositórios digitais confiáveis	USP	Unesp	Unicamp
<u>Integridade</u>			
Desenvolver e adaptar medidas para verificar os dados armazenados e detectar corrupção ou perda. (critérios 8.2.2, 8.2.6, A3.8, B4.4, C1.5, C1.6)	A biblioteca de cada unidade compara a cópia impressa com a digital.	Instâncias administrativas comparam o trabalho depositado com as informações fornecidas pelo autor.	Verificação pré-depósito para conferir a compatibilidade do arquivo digital com a versão impressa do trabalho.
<u>Autenticidade</u>			
Garantir a autenticidade dos documentos (B1.3)	A seção de pós-graduação e a biblioteca da unidade são responsáveis por verificar e validar as informações sobre o trabalho a ser depositado na biblioteca digital.	Considera o autodepósito a garantia da autenticidade, porque as informações são fornecidas pelo autor.	A homologação por parte das secretarias de pós-graduação e da Pró-Reitoria de Pós-Graduação é considerada suficiente para atestar a autenticidade dos trabalhos.
USP e Unicamp consideram a disponibilidade do trabalho na biblioteca digital o certificado de confiabilidade.			
<u>Responsabilidade administrativa</u>			
Registrar políticas e procedimentos e prever revisão e atualização. Comprometer-se com a transparência (critérios A1.1, A3.2, A3.7, B3.1, C3.4)	Não tem política formal, mas registra normas e resoluções, formulários e instruções aos usuários. Não menciona estudos para desenvolver uma política.	Não tem política formal, mas registra instruções de trabalho, formulários e instruções aos usuários. Demonstra intenção de desenvolver uma política.	Não tem política formal, mas registra normas e resoluções, formulários e instruções aos usuários. Demonstra intenção de desenvolver uma política. Tem estudos em andamento.



As três universidades têm prestígio no meio acadêmico e perante a sociedade, elas são selos de garantia para seus projetos e produção intelectual, por isso dificilmente a qualidade deles é questionada. Essas instituições conquistaram respeito. Porém, sob a ótica dos parâmetros para repositórios digitais, as bibliotecas digitais de teses e dissertações apresentadas precisam de melhorias para que sejam efetivamente confiáveis e tenham sucesso na preservação a longo prazo. Os cuidados com a preservação da integridade dos documentos é um aspecto que precisa de mais atenção.

A Unesp e a Unicamp demonstram caminhar para uma importante mudança, ao manifestarem a intenção de formalizar uma política de preservação. Instituições responsáveis pela conservação de um patrimônio precisam ser transparentes, expor suas políticas, seu comprometimento e as ações de preservação adotadas. A transparência conscientiza os usuários sobre a importância da questão, e usuários conscientes serão exigentes e também contribuirão para a manutenção, no caso das bibliotecas deste estudo, ao criarem documentos seguindo padrões recomendados, lembrando que é nessa etapa que devem começar os cuidados para a preservação.

O depósito de um documento no acervo das bibliotecas digitais depende de muitas pessoas para a conferência de informações e inclusão de metadados, e o que se recomenda é ter a maior quantidade possível de metadados gerada automaticamente. Em um ambiente como as universidades, em que existe controle da produção de conhecimento, ou seja, onde se sabe quem é responsável por um projeto de pesquisa e seus produtos (experimentos, produção artística, relatórios, dissertações e teses, artigos científicos, entre outros), parece ser fácil estabelecer estratégias para verificar a autenticidade dos documentos digitais produzidos<sup>105</sup>.

Preservar e verificar a integridade de documentos digitais ainda desperta dúvidas e dificuldades. A USP, Unesp e Unicamp ainda não trabalham muito bem essa questão, nenhuma delas considera a possibilidade de os arquivos sofrerem alterações não intencionais (corrupção ou perda de dados) durante procedimentos de preservação, como a migração. Elas consideram desnecessário verificar a preservação da integridade após o depósito, porque existe o controle de acesso aos arquivos e metadados. O trabalho colaborativo e o intercâmbio de experiências podem contribuir para que as instituições adotem estratégias melhores. Estudar os conceitos de autenticidade e integridade, adotar definições e formalizar o

---

<sup>105</sup> Verificar a autoria, a data de criação e outras informações relevantes para atestar a autenticidade de um documento produzido há anos na universidade, sem testemunhas sobre sua produção, sem ter contato com o próprio autor nem com colegas de trabalho, pode ser mais complicado do que atestar a autenticidade de um documento recentemente produzido, cujo autor está acessível para fornecer informações necessárias.

entendimento da instituição sobre o tema é útil para orientar a criação de políticas, a adoção de estratégias e procedimentos de trabalho<sup>106</sup>.

Por não terem uma política formal para preservação, as universidades não estabeleceram a periodicidade para avaliar as estratégias e os procedimentos, nem a equipe responsável por essa avaliação, mesmo assim, quando necessário, são feitas modificações. Qualquer instituição deveria divulgar atestados ou informações sobre a preocupação em verificar e preservar a autenticidade e a integridade dos documentos digitais sob sua custódia, e, no geral, nenhuma instituição guardiã parece ter determinado uma estratégia de certificação. As práticas das universidades para a preservação de seus acervos digitais não são transparentes; o que encontramos nos respectivos *sites* são normas, principalmente a de criação da biblioteca, e orientações para os autores sobre o depósito. Se existem informações sobre suas práticas, elas precisam ser divulgadas com destaque. No relatório de avaliação produzido pelo CRL, uma das observações mais frequentes é a insuficiência de documentação.

As vantagens que percebemos na atuação das três universidades paulistas são a aparente boa infraestrutura tecnológica, a adoção de padrões e estratégias de preservação amplamente recomendados (*backup* e migração) e a parceria com iniciativas internacionais. A cooperação é a melhor alternativa para desenvolver e aprimorar as práticas.

---

<sup>106</sup> A não adoção de definições para esses conceitos e a não formalização do entendimento sobre o tema podem explicar a falta de ação e a confusão entre a verificação da integridade e da autenticidade. A Unesp, apesar de ser a única universidade que definiu os conceitos, em suas respostas, confundiu estratégias de autenticidade para verificar a integridade.

## 7 CONSIDERAÇÕES FINAIS

O conceito de biblioteca digital ainda não está consolidado, mas percebe-se a preferência por esse termo para se referir a uma coleção ou repositório que compartilhe a missão social e cultural das bibliotecas analógicas e cujo objetivo seja proporcionar acesso a um patrimônio digital e colaborar para o ciclo de criação, disseminação, uso e preservação.

Todas as instituições tradicionalmente responsáveis pela conservação do patrimônio ainda desfrutam respeito e credibilidade conquistados perante a sociedade e isso não mudará, mas precisam se mostrar capazes de enfrentar os desafios impostos pelos avanços tecnológicos e pelas demandas dos usuários. Elas perderam a exclusividade da tarefa de armazenamento e preservação do patrimônio que hoje é compartilhada com terceiros que têm recursos e assumem tal responsabilidade. Vale esclarecer que não se trata de uma disputa, mas de mudanças positivas. É bom que a manutenção do patrimônio não seja responsabilidade de *um*, espera-se que bibliotecas e arquivos tomem a iniciativa, colaborem fortemente, norteiem as discussões e práticas para enfrentar o grande desafio: preservar a longo prazo o patrimônio digital, garantindo não só sua acessibilidade, mas também sua autenticidade e integridade.

Não existe ainda consenso sobre a melhor estratégia de preservação, existem, sim, muitas sugestões e relatos de instituições estrangeiras que estão experimentando propostas, analisando os resultados obtidos, as dificuldades e soluções adotadas. No Brasil, aparentemente faltam relatos do gênero. A maioria deles descreve projetos de digitalização, criação de sites ou repositórios para possibilitar o acesso ao patrimônio, o que parece ser a prioridade, ou seja, os relatos não costumam detalhar a estrutura, as estratégias de preservação adotadas nem as rotinas do gerenciamento.

O rápido avanço tecnológico exige que estratégias de preservação sejam postas em prática o quanto antes, e, apesar da importância, discussões sobre preservação digital e, em especial, sobre conservação da autenticidade e da integridade de documentos parecem ser sempre postergadas. Os projetos não costumam estabelecer políticas antes de entrar em operação, mas preocupam-se com a adoção de padrões que garantam a interoperabilidade e o atendimento satisfatório às demandas dos usuários e que auxiliem na preservação.

Os projetos de biblioteca digital de teses e dissertações da USP, Unesp e Unicamp estão bem encaminhados e têm reconhecimento. Seguem padrões internacionais importantes e atendem a alguns dos critérios propostos para auditoria de repositórios, mas a dedicação à

preservação de seus acervos ainda precisa ser formalizada. Seria interessante realizar uma autoavaliação baseada nos critérios sugeridos pelo Conselho Nacional de Arquivos e organizações internacionais.

Este trabalho não tem a pretensão de julgar as práticas das três universidades estaduais paulistas nem de nenhuma outra instituição que ainda esteja trabalhando nas mesmas condições. Tampouco tem a pretensão de apontar soluções. Não é fácil tomar decisões cujas consequências possam ser irreversíveis, sobretudo quando é necessário considerar aspectos tecnológicos em constante mutação. O objetivo é chamar a atenção para a urgência em sair do campo teórico e formalizar e começar a implantar algumas estratégias. O trabalho cooperativo e a troca de experiências são essenciais, mesmo que a realidade de cada instituição seja diferente e que o maior empecilho provavelmente seja a falta de recursos para garantir a sustentabilidade do projeto de preservação, especialmente a falta de recursos financeiros. Essa é uma das razões pelas quais a preservação digital não pode ser delegada a pessoas físicas ou a pequenas instituições.

Por fim, reforçamos que uma atuação transparente representa grande vantagem e aumenta a credibilidade da instituição perante usuários, parceiros e depositantes. A instituição tem de documentar suas políticas, planos e ações; submeter-se a avaliações periódicas e estar preparada para alterar suas políticas e procedimentos sempre que necessário.

É importante conhecer detalhadamente a prática das instituições nacionais, acadêmicas e de outros segmentos, para traçar um panorama nacional e, então, respeitando as diferenças que provavelmente serão encontradas e as especificidades, criar frentes de atuação coordenadas para avançarmos no processo de preservação do nosso patrimônio digital e buscar garantir sua longevidade, acessibilidade, autenticidade e integridade.

## REFERÊNCIAS

ALMEIDA, Maurício Barcellos. Uma introdução ao XML, sua utilização na internet e alguns conceitos complementares. **Ciência da Informação**, Brasília, v. 31, n. 2, p. 5-13, maio/ago. 2002.

AMERICAN LIBRARY ASSOCIATION. **Definitions of digital preservation**. 2007. Disponível em: <<http://www.lita.org/ala/mgrps/divs/alcts/resources/preserv/defdigpres0408.cfm>>. Acesso em: 23 nov. 2009.

BEARMAN, David; TRANT, Jennifer. Authenticity of digital resources: towards a statement of requirements in the research process. **D-Lib Magazine**, v. 4, n. 6, june 1998. Disponível em: <<http://www.dlib.org/dlib/june98/06bearman.html>>. Acesso em: 2 mar. 2009.

BELLOTTO, Heloísa Liberalli; CAMARGO, Ana Maria de Almeida (Coord.). **Dicionário de terminologia arquivística**. São Paulo: Associação dos Arquivistas Brasileiros – Núcleo Regional de São Paulo; Secretaria de Estado da Cultura, 1996.

BELLOTTO, Heloísa Liberalli. **Arquivística: objetos, princípios e rumos**. São Paulo: Associação de Arquivistas de São Paulo, 2002.

BRAND, Stewart; SANDERS, Terry. Escaping the digital dark age. **Library Journal**, New York, v. 124, n. 2, p. 46-48, 1 fev. 1999.

BUCKLAND, Michael. **What is a “digital document”?**. 1998. Disponível em: <<http://people.ischool.berkeley.edu/~buckland/digdoc.html>>. Acesso em: 2 jul. 2008.

BULLOCK, Alison. Preservation of digital information: issues and current status. **Network Notes**, Ottawa, n. 60, apr. 1999. Disponível em: <<http://epe.lac-bac.gc.ca/100/202/301/netnotes/netnotes-h/notes60.htm>>. Acesso em: 20 mar. 2007.

BUSH, Vannevar. As we may think. **Atlantic Magazine**, july 1945. Disponível em: <<http://www.theatlantic.com/doc/194507/bush>>. Acesso em: 17 abr. 2008.

CABRAL, Maria Luísa. Património cultural e políticas de preservação. In: \_\_\_\_\_. **Amanhã é sempre longe demais: crônicas de P & C**. Lisboa: Gabinete de Estudos, 2002. p. 63-72.

CANFORA, Luciano. As bibliotecas antigas e a história dos textos. In: BARATIN, Marc; JACOB, Christian (Dir.). **O poder das bibliotecas: a memória dos livros no Ocidente**. 2. ed. Rio de Janeiro: UFRJ, 2006. p. 234-245.

CATARINO, Maria Elisabete; BAPTISTA, Ana Alice. Folksonomia: um novo conceito para a organização dos recursos digitais na Web. **DataGramZero**: Revista de Ciência da Informação, Rio de Janeiro, v. 8, n. 3, artigo 4, jun. 2007. Disponível em: <[http://www.datagramazero.org.br/jun07/F\\_I\\_art.htm](http://www.datagramazero.org.br/jun07/F_I_art.htm)>. Acesso em: 15 fev. 2010.

CAVALLO, Guglielmo; CHARTIER, Roger (Org.). Introdução. In: \_\_\_\_\_. **História da leitura no mundo ocidental**. São Paulo: Ática, 1998. v. 1. p. 5-40.

CENTER FOR RESEARCH LIBRARIES. **Ten core requirements**. Jan. 2007. Disponível em: <<http://www.crl.edu/pt-br/archiving-preservation/digital-archives/metrics-assessing-and-certifying/core-re>>. Acesso em: 01 fev. 2010.

CENTRO DE ESTUDOS SOBRE AS TECNOLOGIAS DA INFORMAÇÃO E DA COMUNICAÇÃO. **C14 – Proporção de indivíduos que usam a internet para educação**. Set.-nov. 2009a. Disponível em: <<http://www.cetic.br/usuarios/tic/2009-total-brasil/rel-int-14.htm>>. Acesso em 28 maio 2010.

CENTRO DE ESTUDOS SOBRE AS TECNOLOGIAS DA INFORMAÇÃO E DA COMUNICAÇÃO. **C15 - Atividades desenvolvidas na internet - treinamento e educação**. Set.-nov. 2009b. Disponível em: <<http://www.cetic.br/usuarios/tic/2009-total-brasil/rel-int-15.htm>>. Acesso em 28 maio 2010.

CHAGAS, Mário. Cultura, patrimônio e memória. In: INTEGRAR – CONGRESSO INTERNACIONAL DE ARQUIVOS, BIBLIOTECAS, CENTROS DE DOCUMENTAÇÃO E MUSEUS, 1., 2002, São Paulo. **Textos...** São Paulo: Imprensa Oficial, 2002. p. 135-150.  
CONSELHO NACIONAL DE ARQUIVOS. **Carta para a preservação do patrimônio digital arquivístico**: preservar para garantir o acesso. 2004. Disponível em: <<http://www.documentoseletronicos.arquivonacional.gov.br/media/publicacoes/cartapreservatrimarqdigitalconarq2004.pdf>>. Acesso em: 27 abr. 2007.

CONSELHO NACIONAL DE ARQUIVOS. Câmara Técnica de Documentos Eletrônicos. **Modelo de requisitos para sistemas informatizados de gestão arquivística de documentos**: e-ARQ Brasil. Dez. 2009. Versão 1.1. Disponível em: <<http://www.conarq.arquivonacional.gov.br/media/publicacoes/earqmet/earqbrasilv1.1.pdf>>. Acesso em: 31 maio 2010.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. **Política Nacional de Memória da Ciência e da Tecnologia**: relatório da Comissão Especial constituída pela Portaria 116/2003 do Presidente do CNPq em 04 de julho de 2003. Set. 2003. Disponível em: <<http://ghc.ifi.unicamp.br/SBHC/Memoria-CT.pdf>>. Acesso em: 17 maio 2010.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS. **Reference Model for an Open Archival Information System (OAIS)**. Jan. 2002. Disponível em: <<http://public.ccsds.org/publications/archive/650x0b1.pdf>>. Acesso em: 20 dez. 2009.

CONWAY, Paul. **Preservação no universo digital**. 2. ed. Rio de Janeiro: Projeto Conservação Preventiva em Bibliotecas e Arquivos: Arquivo Nacional, 2001. (Conservação Preventiva em Bibliotecas e Arquivos, 52). Disponível em: <<http://www.arqsp.org.br/cpba/>>. Acesso em: 15 mar. 2007.

CULLEN, Charles T. Authentication of digital objects: lessons from a historian's research. In: COUNCIL ON LIBRARY AND INFORMATION RESOURCES. **Authenticity in a digital environment**. May 2000. p. 1-7. Disponível em: <<http://www.clir.org/PUBS/reports/pub92/pub92.pdf>>. Acesso em: 8 abr. 2007.

CUNHA, Murilo Bastos da. Desafios na construção de uma biblioteca digital. **Ciência da Informação**, Brasília, v. 28, n. 3, p. 257-268, set./dez. 1999. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/285/252>>. Acesso em: 15 mar. 2007.

CUNHA, Murilo Bastos da; MCCARTHY, Cavan. Estado atual das bibliotecas digitais no Brasil. In: MARCONDES, Carlos H. et al. (Org.). **Bibliotecas digitais: saberes e práticas**. 2. ed. Salvador: UFBA, 2006. p. 25-54.

CUNHA, Murilo Bastos da. **Dicionário de Biblioteconomia e Arquivologia**. Brasília: Briquet de Lemos, 2008.

DigiCULT Project. **Integrity and authenticity of digital cultural heritage objects**. Aug. 2002. (Thematic Issue 1). Disponível em: <[http://www.digicult.info/downloads/thematic\\_issue\\_1\\_final.pdf](http://www.digicult.info/downloads/thematic_issue_1_final.pdf)>. Acesso em: 3 set. 2007.

DIGITAL LIBRARY FEDERATION. **A working definition of digital library**. 1998. Disponível em: <<http://www.diglib.org/about/dldefinition.htm>>. Acesso em: 26 maio 2007.

DUGUID, Paul. **Report of the Santa Fe Planning Workshop on Distributed Knowledge Work Environments: digital libraries**. University of Michigan School of Information, sept. 1997. Disponível em: <<http://web.archive.org/web/19981202153946/http://www.si.umich.edu/SantaFe/>>. Acesso em: 5 maio 2009.

DURANTI, Luciana. Reliability and authenticity: the concepts and their implications. **Archivaria**, n. 39, p. 5-10, spring 1995. Disponível em: <<http://journals.sfu.ca/archivar/index.php/archivaria/article/view/12063/13035>>. Acesso em: 15 mar. 2009.

DURANTI, Luciana. Introducción. In: \_\_\_\_\_ (Ed.). **La conservación a largo plazo de documentos electrónicos auténticos: hallazgos del Proyecto InterPares**. Cartagena: Concejalía de Cultura : 3000 Informática, 2005. p. 15-22.

EDUCOPIA INSTITUTE. MetaArchive Cooperative charter: a charter describing the purposes and aims of the MetaArchive Cooperative, an association dedicated to the preservation of cultural memory materials that are digital in nature and form. Nov. 2009. Disponível em:

<[http://www.metaarchive.org/public/resources/charter\\_member/MetaArchive\\_Charter\\_2010.pdf](http://www.metaarchive.org/public/resources/charter_member/MetaArchive_Charter_2010.pdf)>. Acesso em: 2 jun. 2010.

EUROPEAN TASK FORCE PERMANENT ACCESS. **Permanent access to the records of science**: strategic action programme 2006-2010. Nov. 2005. Disponível em: <<http://www.alliancepermanentaccess.eu/Strategic%20Action%20Programme.pdf>>. Acesso em: 2 jun. 2010.

FERREIRA, Miguel. **Introdução à preservação digital**: conceitos, estratégias e actuais consensos. Guimarães: Escola de Engenharia da Universidade do Minho, 2006. Disponível em: <<http://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>>. Acesso em: 2 abr. 2008.

FOGOLIN, D. F.; KEMPINAS, A. L. G. Implantação da biblioteca digital de teses e dissertações da Universidade Estadual Paulista – Unesp. In: Simpósio Internacional de Bibliotecas Digitais, 2., 2004, Campinas. Disponível em: <<http://libdigi.unicamp.br/document/?code=8267>>. Acesso em: 19 jan. 2010.

GNU Project. **Categories of free and non-free software**. Disponível em: <<http://www.gnu.org/philosophy/categories.html>>. Acesso em: 20 dez. 2009.

GONTIJO, Silvana. **O livro de ouro da comunicação**. Rio de Janeiro: Ediouro, 2004.

GOULEMOT, Jean-Marie. Bibliotecas, enciclopedismo e angústias da perda: a exaustividade ambígua das Luzes. In: BARATIN, Marc; JACOB, Christian (Dir.). **O poder das bibliotecas**: a memória dos livros no Ocidente. 2. ed. Rio de Janeiro: UFRJ, 2006. p. 257-270.

HACKETT, Yvette; UNDERWOOD, William; EPPARD, Philip. Part one – case and general studies in the artistic, scientific and governmental sectors : Focus Task Force Report. In: DURANTI, Luciana; PRESTON, Randy (Ed.). **International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2**: experiential, interactive and dynamic records. Padova: Associazione Nazionale Archivistica Italiana, 2008. Disponível em: [http://www.interpares.org/ip2/display\\_file.cfm?doc=ip2\\_book\\_part\\_1\\_focus\\_task\\_force.pdf](http://www.interpares.org/ip2/display_file.cfm?doc=ip2_book_part_1_focus_task_force.pdf). Acesso em: 15 fev. 2009.

HARTER, Stephen P. Scholarly communication and the digital library: problems and issues. **Journal of Digital Information**, Texas, v. 1, n. 1, 1997. Disponível em: <<http://journals.tdl.org/jodi/article/view/4/4>>. Acesso em: 20 jun. 2009.



HEDSTROM, Margaret. Digital preservation: a time bomb for digital libraries. **Computer and the humanities**, Netherlands, n. 31, p. 189-202, 1998. Disponível em: <[http://deepblue.lib.umich.edu/bitstream/2027.42/42573/1/10579\\_2004\\_Article\\_153071.pdf](http://deepblue.lib.umich.edu/bitstream/2027.42/42573/1/10579_2004_Article_153071.pdf)> Acesso em: 20 maio 2007.

HERÓDOTO. **História**: o relato clássico da guerra entre gregos e persas. 2. ed. São Paulo: Prestígio, 2001.

IBGE INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Tabela 1.22.1 - Pessoas de 10 anos ou mais de idade que utilizaram a Internet, no período de referência dos últimos três meses, por Grandes Regiões, segundo o local de acesso à Internet e os grupos de anos de estudo**. 2005a. Disponível em: <[http://www.ibge.gov.br/home/estatistica/populacao/acesoainternet/tabelas/tab1\\_22\\_1.pdf](http://www.ibge.gov.br/home/estatistica/populacao/acesoainternet/tabelas/tab1_22_1.pdf)>. Acesso em: 15 jul. 2009.

IBGE INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Tabela 1.40.2 - Distribuição das pessoas de 10 anos ou mais de idade que utilizaram a Internet no domicílio em que moravam, no período de referência dos últimos três meses, total e que utilizaram a Internet somente no domicílio em que moravam, por Grandes Regiões, segundo o tipo de conexão à Internet no domicílio**. 2005b. Disponível em: <[http://www.ibge.gov.br/home/estatistica/populacao/acesoainternet/tabelas/tab1\\_40\\_2.pdf](http://www.ibge.gov.br/home/estatistica/populacao/acesoainternet/tabelas/tab1_40_2.pdf)>. Acesso em: 15 jul. 2009.

IBGE INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Tabela 1.26.2 - Percentual das pessoas que utilizaram a Internet para cada finalidade, na população de 10 anos ou mais de idade que utilizou a Internet, no período de referência dos últimos três meses, por Grandes Regiões e sexo, segundo a finalidade do acesso à Internet**. 2005c. Disponível em: <[http://www.ibge.gov.br/home/estatistica/populacao/acesoainternet/tabelas/tab1\\_26\\_2.pdf](http://www.ibge.gov.br/home/estatistica/populacao/acesoainternet/tabelas/tab1_26_2.pdf)>. Acesso em: 15 jul. 2009.

IBOPE. **Internet no Brasil cresceu 10% no mês de julho**. Ago. 2009. Disponível em: <[http://www.ibope.com.br/calandraWeb/servlet/CalandraRedirect?temp=6&proj=PortalIBOPE&pub=T&db=caldb&comp=pesquisa\\_leitura&nivel=null&docid=62A33B253477B58783257619004BD15C](http://www.ibope.com.br/calandraWeb/servlet/CalandraRedirect?temp=6&proj=PortalIBOPE&pub=T&db=caldb&comp=pesquisa_leitura&nivel=null&docid=62A33B253477B58783257619004BD15C)>. Acesso em: 17 jan. 2010.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS. **Functional Requirements for Bibliographic Records**: final report – current text. [2009]. cap. 2. Disponível em: <[http://archive.ifla.org/VII/s13/frbr/frbr\\_current2.htm](http://archive.ifla.org/VII/s13/frbr/frbr_current2.htm)>. Acesso em: 10 maio 2010.

InterPARES. **Project overview**. Disponível em: <<http://www.interpares.org/welcome.cfm>>. Acesso em: 15 fev. 2009.

INTO THE FUTURE: on the preservation of knowledge in the electronic age. Direção: Terry Sanders. Narração: Robert MacNeil. Santa Monica: American Film Foundation, 1998. 1 videocassete (33 min), DVD, son., color.

KONDO, Rogério Toshiaki; LIRANI, Maria de Lourdes Rebucci; TRAINA JR., Caetano. Bibliotecas digitais: a experiência da USP. **Revista USP**, n. 80, p. 62-71, dez./fev. 2008-2009.

ISO INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 15489-1**: information and documentation: records management. Set. 2009. Disponível em: <[http://www.javeriana.edu.co/archivo/07\\_eventos/preservaciondigital/memorias/index\\_archivos/norma/iso\\_15489-1.pdf](http://www.javeriana.edu.co/archivo/07_eventos/preservaciondigital/memorias/index_archivos/norma/iso_15489-1.pdf)>. Acesso em: 17 maio 2007.

KUNY, Terry; CLEVELAND, Gary. The digital library: myths and challenges. **IFLA Journal**, v. 24, n. 2, p. 107-113, 1998. Disponível em: <<http://archive.ifla.org/V/iflaj/jour2402.pdf>>. Acesso em: 30 jul. 2009.

KURAMOTO, Hélio. Biblioteca digital brasileira: integrando a ICT brasileira. In: MARCONDES, Carlos H. et al. (Org.). **Bibliotecas digitais: saberes e práticas**. 2. ed. Salvador: UFBA, 2006. p. 287-303.

LE COADIC, Yves-François. **A ciência da informação**. 2. ed. rev. ampl. Brasília: Brinquet de Lemos, 2004.

LE CROSNIER, Hervé. Bibliotecas digitais. 27 mar 2006. In: AMBROSI, Alain; PEUGEOT, Valérie; PIMIENTA, Daniel (Coord.). **Desafios de palavras: enfoques multiculturais sobre as sociedades da informação**. Caen: C & F éditions, 2005. Disponível em: <<http://www.vecam.org/article628.html>>. Acesso em: 22 maio 2007.

LEVACOV, Marília. Biblioteca virtuais: (r)evolução?. **Ciência da Informação**, Brasília, v. 26, n. 2, p. 125-135, ago. 1997. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/390/350>>. Acesso em: 15 mar. 2007.

LEVY, David M. Where's Waldo? Reflections on copies and authenticity in a digital environment. In: COUNCIL ON LIBRARY AND INFORMATION RESOURCES. **Authenticity in a digital environment**. May 2000. p. 24-31. Disponível em: <<http://www.clir.org/PUBS/reports/pub92/pub92.pdf>>. Acesso em: 8 abr. 2007.

LIBRARY OF CONGRESS. **Understanding MARC**. Disponível em: <<http://www.loc.gov/marc/umb/um01to06.html>>. Acesso em: 28 dez. 2009.

LORIST, H. H. J.; MEER, K. van der. **Standards for digital libraries and archives: digital longevity**. [2001?]. Disponível em: <<http://faculty.fims.uwo.ca/nickerson/520/Reserves/Standards%20and%20Longevity%20with%20Digital%20Libraries.pdf>>. Acesso em: 25 dez. 2009.

LYNCH, Clifford. Canonicalization: a fundamental tool to facilitate preservation and management of digital information. **D-Lib Magazine**, v. 5, n. 9, sept. 1999. Disponível em: <<http://www.dlib.org/dlib/september99/09lynch.html>>. Acesso em: 20 jan. 2010.

LYNCH, Clifford. Authenticity and integrity in the digital environment: an exploratory analysis of the central role of trust. In: COUNCIL ON LIBRARY AND INFORMATION RESOURCES. **Authenticity in a digital environment**. May 2000. p. 32-50. Disponível em: <<http://www.clir.org/PUBS/reports/pub92/pub92.pdf>>. Acesso em: 8 abr. 2007.

MACNEIL Heather; GILLILAND-SWETLAND, Anne. Informe del grupo de trabajo sobre autenticidad. In: DURANTI, Luciana (Ed.). **La conservación a largo plazo de documentos electrónicos auténticos: hallazgos del Proyecto InterPares**. Cartagena: Concejalía de Cultura: 3000 Informática, 2005a. p. 25-82.

[MACNEIL Heather; GILLILAND-SWETLAND, Anne]. Apéndice 2: requisitos para ponderar y mantener la autenticidad de los documentos electrónicos. In: DURANTI, Luciana (Ed.). **La conservación a largo plazo de documentos electrónicos auténticos: hallazgos del Proyecto InterPares**. Cartagena: Concejalía de Cultura : 3000 Informática, 2005b. p. 259-276.

MACNEIL, Heather; MAK, Bonnie. Constructions of authenticity. **Library Trends**, Baltimore, Maryland, v. 56, n. 1, summer 2007, p. 26-52.

MARQUES, Eliana de Azevedo. A nova biblioteca: o papel e o digital. **Revista USP**, n. 80, p. 18-27, dez./fev. 2008-2009.

MARTINS, Roberto de A. A memória científica nas universidades: estratégias para preservação do patrimônio científico e tecnológico brasileiro. Disponível em: <<http://ghtc.ifi.unicamp.br/pdf/ram-121.pdf>>. Acesso em: 10 abr. 2010.

MARTINS, Wilson. **A palavra escrita: história do livro, da imprensa e da biblioteca**. 3. ed. rev. e atual. São Paulo: Ática, 2002.

MAXIMIANO, Antonio Cesar Amaru. **Introdução à administração**. 5. ed. rev. ampl. São Paulo: Atlas, 2000.

MCMILLAN, Gail; SKINNER, Katherine. **NDLTD Preservation Strategy with the MetaArchive Cooperative**. 2nd ed. June 2009. Disponível em: <<http://scholar.lib.vt.edu/theses/NDLTD/BoD200906/NDLTDPreservationPlan20090606.pdf>>. Acesso em: 2 jun. 2010.

MCPHERSON, Madelaine. Managing digital libraries. In: CSIRO INFORMATION, MANAGEMENT & TECHNOLOGY CONFERENCE, 1997, Gold Coast. Disponível em: <<http://www.usq.edu.au/users/mcpherso/csiro.htm>>. Acesso em: 18 jul. 2009.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **Understanding metadata**. 2004. Disponível em: <<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>>. Acesso em: 27 dez. 2009.

ONLINE COMPUTER LIBRARY CENTER. **Preservation metadata and the OAIS information model**: a metadata framework to support the preservation of digital objects. June 2002. Disponível em:

<[http://www.oclc.org/research/activities/past/orprojects/pmwg/pm\\_framework.pdf](http://www.oclc.org/research/activities/past/orprojects/pmwg/pm_framework.pdf)>. Acesso em: 22 nov. 2009.

ONLINE COMPUTER LIBRARY CENTER; CENTER FOR RESEARCH LIBRARIES. **Trustworthy repositories audit & certification**: criteria and checklist. Version 1.0. Feb. 2007. Disponível em:

<[http://www.crl.edu/sites/default/files/attachments/pages/trac\\_0.pdf](http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf)>. Acesso em: 1 fev. 2010.

OPEN ARCHIVES INITIATIVE. **The Open Archives Initiative**: mission statement. Disponível em: <<http://www.openarchives.org/OAI/OAI-organization.php>>. Acesso em: 17 jan. 2010.

OWEN, John Mackenzie. Preserving the digital heritage: roles and responsibilities for heritage repositories. In: LUSENET, Yola de; WINTERMANS, Vincent (Ed.). **Preserving the digital heritage**: principles and policies. 2007. p. 45-49. Disponível em: <<http://www.knaw.nl/ecpa/publ/pdf/2735.pdf>>. Acesso em: 11 nov. 2008.

PEREIRA, Edmeire Cristina; RUTINA, Raquel. O século XXI e o sonho da biblioteca universal: quase seis mil anos de evolução na produção, registro e socialização do conhecimento. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 4, n. 1, p. 5-19, jan./jun. 1999. Disponível em:

<<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/viewFile/590/359>>. Acesso em: 24 maio 2007.

PUNTONI, Pedro. As bibliotecas digitais e a sociedade da informação: perspectivas para as bibliotecas digitais no Brasil. **Revista USP**, n. 80, p. 44-53, dez./fev. 2008-2009.

RAYWARD, W. Boyd. The origins of information science and the International Institute of Bibliography/ International Federation for Information and Documentation (FID). **Journal of the American Society for Information Science**, v. 48, n. 4, p. 289-300, 1997.

REAGAN, Brad. The digital ice age. **Popular Mechanics**, New York, oct. 2009. Section Technology : Gadgets : Industry News. Disponível em:

<<http://www.popularmechanics.com/technology/industry/4201645.html?page=1>>. Acesso em: 14 fev. 2007.

REGISTRY OF OPEN ACCESS REPOSITORIES. **Browse by country**. Disponível em: <<http://roar.eprints.org/view/geoname/>>. Acesso em: 14 fev. 2010.

RESEARCH LIBRARIES GROUP. **Trusted digital repositories**: attributes and responsibilities. May 2002. Disponível em: <<http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf>>. Acesso em: 22 nov. 2009.

RIVERA DONOSO, Miguel Ángel. Directrices para la creación de un programa de preservación digital. **Serie Bibliotecología y Gestión de Información**, n. 43, marzo 2009. Disponível em: <[http://eprints.rclis.org/16052/1/Serie\\_N%C2%B043\\_Preservacion\\_digital.pdf](http://eprints.rclis.org/16052/1/Serie_N%C2%B043_Preservacion_digital.pdf)>. Acesso em: 30 set. 2009.

ROEDER, John; EPPARD, Philip; UNDERWOOD, William; LAURIAULT, Tracey P. Part three - Authenticity, reliability and accuracy of digital records in artistic, scientific and governmental sectors : Domain 2 task force Report. In: DURANTI, Luciana; PRESTON, Randy (Ed.). **International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2**: experiential, interactive and dynamic records. Padova: Associazione Nazionale Archivistica Italiana, 2008. Disponível em: <[http://www.interpares.org/ip2/display\\_file.cfm?doc=ip2\\_book\\_part\\_3\\_domain2\\_task\\_force.pdf](http://www.interpares.org/ip2/display_file.cfm?doc=ip2_book_part_3_domain2_task_force.pdf)>. Acesso em: 15 fev. 2009.

ROTHENBERG, Jeff. **Avoiding technological quicksand**: finding a viable technical foundation for digital preservation. Jan. 1999. Disponível em: <<http://www.clir.org/pubs/reports/rothenberg/pub77.PDF>>. Acesso em: 13 mar. 2007.

ROTHENBERG, Jeff. **Ensuring the longevity of digital information**. Feb. 1999. Disponível em: <<http://www.clir.org/pubs/archives/ensuring.pdf>>. Acesso em: 13 mar. 2007.

ROTHENBERG, Jeff. Preserving authentic digital information. In: COUNCIL ON LIBRARY AND INFORMATION RESOURCES. **Authenticity in a digital environment**. May 2000. p. 51-68. Disponível em: <<http://www.clir.org/PUBS/reports/pub92/pub92.pdf>>. Acesso em: 8 abr. 2007.

ROWLEY, Jennifer. **A biblioteca eletrônica**. Brasília: Briquet de Lemos, 2002.

SANTOS, Vanderlei Batista dos. **Gestão de documentos eletrônicos: uma visão arquivística**. 2. ed. rev. e aum. Brasília: ABARQ, 2005.

SARACEVIC, Tefko. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, v. 1, n. 1, p. 41-62, jan./jun. 1996. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/viewFile/235/22>>. Acesso em: 8 maio 2008.

SAYÃO, Luís Fernando. Preservação digital no contexto das bibliotecas digitais: uma breve introdução. In: MARCONDES, Carlos H. et al. (Org.). **Bibliotecas digitais: saberes e práticas**. 2. ed. Salvador: UFBA, 2006. p. 113-143.

SAYÃO, Luís Fernando. Interoperabilidade das bibliotecas digitais: o papel dos sistemas de identificadores persistentes – URN, PURL, DOI, Handle System, CrossRef e OpenURL. **TransInformação**, Campinas, v. 19, n. 1, p. 65-82, jan./abr. 2007.

SAYÃO, Luís Fernando. Bibliotecas digitais e suas utopias. **Ponto de Acesso**, Salvador, v. 2, n. 2, p. 2-36, ago./set. 2008. Disponível em: <[www.portalseer.ufba.br/index.php/revistaici/article/view/2661/2166](http://www.portalseer.ufba.br/index.php/revistaici/article/view/2661/2166)>. Acesso em: 15 maio 2009.

SAYÃO, Luis Fernando. Afinal, o que é biblioteca digital?. **Revista USP**, n. 80, p. 6-17, dez./fev. 2008-2009.

SBU SISTEMA DE BIBLIOTECAS DA UNICAMP. **Biblioteca digital da Unicamp**: apresentação. Disponível em: <<http://www.sbu.unicamp.br/bibdig/apresentacao.htm>>. Acesso em: 10 jan. 2010.

SILVA, Helena Pereira da; JAMBEIRO, Othon; BARRETO, Ângela Maria. Bibliotecas digitais: uma nova cultura, um novo conceito, um novo profissional. In: MARCONDES, Carlos H. et al. (Org.). **Bibliotecas digitais**: saberes e práticas. 2. ed. Salvador: UFBA, 2006. p. 259-284.

SKINNER, Katherine; MEVENKAMP, Monika. DDP Architecture. In: SKINNER, Katherine; SCHULTZ, Matt (Ed.). **A guide to distributed digital preservation**. Atlanta, GA: Educopia, 2010. p. 11-25. Disponível em: <[http://www.metaarchive.org/sites/default/files/GDDP\\_Educopia.pdf](http://www.metaarchive.org/sites/default/files/GDDP_Educopia.pdf)>. Acesso em: 5 jun. 2010.

SMITH, Abby. Introduction. In: COUNCIL ON LIBRARY AND INFORMATION RESOURCES. **Authenticity in a digital environment**. May 2000a. p. vi-viii. Disponível em: <<http://www.clir.org/PUBS/reports/pub92/pub92.pdf>>. Acesso em: 8 abr. 2007.

SMITH, Abby. Authenticity in perspective. In: COUNCIL ON LIBRARY AND INFORMATION RESOURCES. **Authenticity in a digital environment**. May 2000b. p. 69-75. Disponível em: <<http://www.clir.org/PUBS/reports/pub92/pub92.pdf>>. Acesso em: 8 abr. 2007.

SOUZA, Márcia Izabel F.; VENDRUSCULO, Laurimar G.; MELO, Geane Cristina. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. **Ciência da Informação**, Brasília, v. 29, n. 1, p. 93-102, jan./abr. 2000.

SUGIMOTO, Luiz. **Biblioteca digital publica tese 30 mil e dá acesso a 100% do acervo da Unicamp**. Out. 2009. Disponível em: <<http://www.unicamp.br/unicamp/divulgacao/2009/10/14/biblioteca-digital-publica-tese-30000-nil-e-da-acesso-a-100-do-acervo-da-unica>>. Acesso em: 10 jan. 2010.

TAMMARO, Anna Maria. Conversações sobre a biblioteca digital. In: TAMMARO, Anna Maria; SALARELLI, Alberto. **A biblioteca digital**. Brasília: Briquet de Lemos, 2008. p. 111-142.

TANSELLE, G. Thomas. Thoughts on the authenticity of electronic texts. **Studies in Bibliography**, v. 54, p. 133-136, 2001. Disponível em: <<http://etext.virginia.edu/bsuva/sb/>>. Acesso em: 18 jul. 2009.

TASK FORCE ON ARCHIVING OF DIGITAL INFORMATION. **Preserving digital information**. May 1996. Disponível em: <<http://www.oclc.org/research/activities/past/rlg/digpresstudy/final-report.pdf>>. Acesso em: 13 mar. 2007.

TENNANT, Roy. **Digital v. electronic v. virtual libraries**. 1999. Disponível em: <<http://sunsite.berkeley.edu/mydefinitions.html>>. Acesso em: 8 jul. 2009.

UNESCO UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION. **Charter on the preservation of the digital heritage**. Oct. 2003. Disponível em: <[http://portal.unesco.org/en/ev.php-URL\\_ID=17721&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/en/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html)>. Acesso em: 20 mar. 2007.

WORLD WIDE WEB CONSORTIUM. **XML essentials**. Disponível em: <<http://www.w3.org/standards/xml/core>>. Acesso em: 20 dez. 2009.

WALLIS, Jillian C. et al. **Know thy sensor**: trust, data quality, and data integrity in scientific digital libraries. Aug. 2007. Disponível em: <<http://www.escholarship.org/uc/item/4xx221vv>>. Acesso em: 3 abr. 2010.

## **ANEXO A**

### **Questionário**



1. Existe uma política especialmente criada para a preservação digital em longo prazo do acervo da biblioteca?
2. Estão determinadas na política as estratégias escolhidas para garantir a preservação do acervo?
  - Caso não exista uma política, ainda assim são adotadas estratégias de preservação? Quais?
3. Existe alguma explicação sobre as estratégias e/ou o porquê de sua escolha?
4. Os conceitos de autenticidade e de integridade de documentos digitais estão presentes na política de preservação?
  - Em caso positivo, quais foram as definições adotadas pela biblioteca para esses dois conceitos? Quais são as estratégias e procedimentos adotados para preservação?
  - Em caso negativo, por quê?
5. Como a biblioteca verifica a autenticidade e a integridade de documento a ser incluído no acervo?
6. Quais são os metadados de preservação adotados? Como eles são gerados?
7. Quais são as estratégias adotadas para controlar esses metadados?
8. Com que frequência se verifica a autenticidade e a integridade dos documentos?
9. Quais são as estratégias adotadas para tal verificação?
10. Como a biblioteca demonstra a seus usuários a autenticidade e a integridade de um documento?
11. Já foi constatado algum tipo de alteração em um documento? Em caso positivo, como a alteração foi detectada? Qual foi o procedimento a partir dessa descoberta?

12. Qual é a periodicidade da avaliação das políticas de preservação? Quem são os responsáveis pela avaliação?

13. As políticas para preservação da autenticidade e da integridade já precisaram sofrer alguma alteração desde sua criação? Em caso positivo, por quê?

14. Como foram registradas essas alterações?

15. As políticas são divulgadas aos usuários? Em caso negativo, por quê? Há intenção de divulgá-las?

Se for possível, favor, fornecer uma cópia da atual política de preservação.

## ANEXO B

### Respostas recebidas da USP<sup>107</sup>

---

<sup>107</sup> Em algumas respostas foram mencionados os procedimentos adotados para a Biblioteca Digital de Obras Raras e Especiais da Universidade (<http://www.obrasraras.usp.br/>).

**1. Existe uma política especialmente criada para a preservação digital em longo prazo do acervo da biblioteca?**

R: A política seguida para as bibliotecas digitais não está formalizada, mas seguem princípios gerais baseados nas recomendações de boas práticas das organizações internacionais pertinentes. Os critérios adotados para a segurança do meio são: esquema de *backup* em disco e cartucho, espelhamento em disco através de sistema de *storage* e, em implementação espelhamento de equipamento de *storage* (um na área 1 e outro na área 2 do campus São Carlos, distante 5 quilômetros um do outro). A atualização do *hardware* e *software* (incluindo formato de arquivos e tipo de mídia) é continuamente avaliada e, quando necessário, mudanças são realizadas no sentido de evitar obsolescência (tanto de *hardware* quanto de *software*). Nesse sentido, iniciamos na plataforma NDLTD (baseado em linguagem *perl*) e estamos hoje em plataforma própria (linguagem *php* e tecnologia AJAX). No decorrer desse período também foi acrescentado o processo de *harvesting* pelo protocolo OAI. Finalmente, a garantia da compatibilidade dos arquivos é assegurada pela análise do padrão dos arquivos submetidos quando de sua inserção nas bibliotecas, o que garante sua compatibilidade com os padrões estabelecidos.

**2. Estão determinadas na política as estratégias escolhidas para garantir a preservação do acervo?**

- Caso não exista uma política, ainda assim são adotadas estratégias de preservação? Quais?

R: Resposta inclusa na questão 1.

**3. Existe alguma explicação sobre as estratégias e/ou o porquê de sua escolha?**

R: Evolução tecnológica, adoção de padrões internacionais e definição de protocolos de boas práticas recomendadas pelos organismos internacionais.

**4. Os conceitos de autenticidade e de integridade de documentos digitais estão presentes na política de preservação?**

- Em caso positivo, quais foram as definições adotadas pela biblioteca para esses dois conceitos? Quais são as estratégias e procedimentos adotados para preservação?

R: Sim. Veja na questão 5.

- Em caso negativo, por quê?

**5. Como a biblioteca verifica a autenticidade e a integridade de documento a ser incluído no acervo?**

R: a) No caso de obras raras, a autenticidade é confirmada na medida que o livro em meio digital é digitalizado pelo próprio Sistema de Bibliotecas da USP e inserido na biblioteca internamente, sem a possibilidade de alterações externas.

b) No caso de teses/dissertações, conforme procedimentos existentes na universidade, os dados são validados pela Seção de Pós-Graduação e a seguir pela biblioteca da unidade. Nesse processo, a seção recebe a cópia impressa, válida e homologa, via sistema, com os dados da cópia digital e dados administrativos.

**6. Quais são os metadados de preservação adotados? Como eles são gerados?**

R: Os metadados bibliográficos, no formato Dublin Core, referentes a obras raras e teses/dissertações, são produzidos por bibliotecários especialistas em catalogação de tipo de publicações. O sistema gera metadados internos para gestão administrativa, em complementação aos dados bibliográfico como a data de publicação do trabalho na internet.

**7. Quais são as estratégias adotadas para controlar esses metadados?**

R: No caso de metadados bibliográficos, somente os profissionais devidamente habilitados no sistema é que têm acesso para a criação, alteração e remoção.

No caso de teses e dissertações, os metadados provêm dos sistemas corporativos da USP.

**8. Com que frequência se verifica a autenticidade e a integridade dos documentos?**

R: Não é necessário. Tendo em vista o fluxo adotado que já estabelece um critério de verificação da integridade. (Neste caso pode-se ver a referência a seguir)

MASIERO, P. C.; BREMER, C. F.; COLETTA, T. G.; LIRANI, M. L. R.; KONDO, R. T.; ARAGÃO, A. C.; MOSCONI, E. P.; SALEM, A. D. C. A Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo. Ciência da Informação, v. 30, p. 34-41, 2001.

Disponível na página do CISC - <http://www.cisc.usp.br> - item: publicações- artigos completos publicados em periódicos.

**9. Quais são as estratégias adotadas para tal verificação?**

R: (Idem)

**10. Como a biblioteca demonstra a seus usuários a autenticidade e a integridade de um documento?**

R: A autenticidade de uma informação sobre o documento é validada na medida em que sua inserção nas base de dados somente é realizada por profissionais devidamente habilitados para a operacionalização dos sistemas envolvidos nesse tipo de atividade.

**11. Já foi constatado algum tipo de alteração em um documento? Em caso positivo, como a alteração foi detectada? Qual foi o procedimento a partir dessa descoberta?**

R: Muito raramente, mas já foi necessária alguma correção no documento de uma tese. O processo é iniciado pela CPG da unidade da defesa da tese, que tem autoridade para efetuar a alteração, mas não os meios para isso. A alteração é realizada pelo CISC após receber a devida autorização dos órgãos competentes. Note-se que a biblioteca é mantida com dois sistemas: o sistema de produção, acessível apenas pelos órgãos internos da USP, e o sistema de consulta, disponível publicamente. Os dados são preparados no sistema de produção, onde é feito o controle de entrada e manutenção. Quando um documento é homologado, ele é copiado para o sistema de consulta. Como não existe acesso público ao sistema de produção, não é possível a adulteração dos originais da tese, que são periodicamente copiados para o sistema de consulta.

**12. Qual é a periodicidade da avaliação das políticas de preservação? Quem são os responsáveis pela avaliação?**

R: A preservação ocorre automaticamente pela interação entre os sistemas de produção, consulta e *backup*.

**13. As políticas para preservação da autenticidade e da integridade já precisaram sofrer alguma alteração desde sua criação? Em caso positivo, por quê?**

R: Uma vez que o processo de criação, validação, homologação e disponibilização pública dos documentos segue as normas da Pró-Reitoria de Pós-graduação, SIBI e das unidades envolvidas, o processo está vinculado às decisões políticas desses organismos, que efetivamente ocorrem e são atendidas pela equipe técnica.

**14. Como foram registradas essas alterações?**

R: Pela publicação de normas e resoluções.

**15. As políticas são divulgadas aos usuários? Em caso negativo, por quê? Há intenção de divulgá-las?**

R: As normas são sempre amplamente divulgadas.

**Se for possível, favor, fornecer uma cópia da atual política de preservação.**

## **ANEXO C**

### **Respostas recebidas da Unesp**



**1. Existe uma política especialmente criada para a preservação digital em longo prazo do acervo da biblioteca?**

R: Não há uma política formalizada, contudo existem ações relacionadas à preservação digital de informações eletrônicas e digitais.

**2. Estão determinadas na política as estratégias escolhidas para garantir a preservação do acervo?**

- Caso não exista uma política, ainda assim são adotadas estratégias de preservação? Quais?

R: Existem ações voltadas para a preservação digital, dentre elas podem-se citar:

- 1) *Upgrade* de versão dos sistemas de informação (*softwares*) utilizados na Rede de Bibliotecas.
- 2) *Upgrade* de formatos (.doc; .pdf; .jpg etc.) de dados/informações armazenados nos sistemas de informação utilizados na Rede de Bibliotecas.
- 3) Migração de dados/informações em plataformas mais modernas (*hardware*).
- 4) Aquisição de arquivos perpétuos de coleções eletrônicas/digitais.
- 5) Aquisição de *backfiles* de coleções eletrônicas/digitais.

**3. Existe alguma explicação sobre as estratégias e/ou o porquê de sua escolha?**

R: A escolha se deve a três fatores:

- 1) Análise das tendências mundiais;
- 2) Análise das ações realizadas pelas grandes universidades de países desenvolvidos;
- 3) Análise das necessidades da Rede de Bibliotecas.

A partir disso, decidimos sobre os *softwares*, *hardwares*, formatos e ações que devem ser implementados/realizados no âmbito institucional.

**4. Os conceitos de autenticidade e de integridade de documentos digitais estão presentes na política de preservação?**

R: Não há uma política de preservação formal na instituição. Contudo, a questão da autenticidade e integridade dos dados/informação é observada, bem como todos os quesitos relacionados aos direitos autorais são monitorados.

- Em caso positivo, quais foram as definições adotadas pela biblioteca para esses dois conceitos? Quais são as estratégias e procedimentos adotados para preservação?

R: Conforme mencionado, não há uma política formal em relação a isso, mas compreendemos autenticidade como “[...] qualidade, condição ou caráter do que é genuíno, verdadeiro [...]”, ou seja, entendemos autenticidade como uma forma de comprovação da autoria do dado/informação, além disso há a total confiança quanto a ausência de falsidade/falsificação. No caso da integridade, a compreendemos como “[...] estado ou característica daquilo que [...] não sofreu nenhuma diminuição; plenitude, inteireza [...] característica ou estado daquilo que se apresenta ileso, intacto, que não foi atingido ou agredido”, isto é, os dados/informações contidos nos documentos/suportes não podem ter sido adulterados após a sua criação/produção, uma vez que serão objetos de prova legal/jurídica. Assim, quando nos referimos à integridade digital, há de ter uma estrutura que garanta a integridade ou que gere evidências das alterações sofridas após a criação/produção.

- Em caso negativo, por quê?

#### **5. Como a biblioteca verifica a autenticidade e a integridade de documento a ser incluído no acervo?**

R: No caso da Rede de Bibliotecas, a autenticidade pode ser verificada pela submissão do documento (teses, dissertações, trabalhos de conclusão de curso), uma vez que o próprio autor submete o documento eletrônico/digital, autorizando sua armazenagem e disseminação. A integridade pode ser verificada com base no preenchimento de formulário específico para esse fim, contendo todas as informações sobre o conteúdo, e assinado pelo próprio autor. Além disso, as instâncias administrativas internas à instituição verificam a correspondência do que foi mencionado e depositado com o documento eletrônico/digital propriamente dito.

#### **6. Quais são os metadados de preservação adotados? Como eles são gerados?**

R: Os metadados são gerados a partir da entrada dos dados no sistema e inseridos de acordo com as normas internacionais do formato MARC, Dublin Core, entre outros.

#### **7. Quais são as estratégias adotadas para controlar esses metadados?**

R: Há um trabalho cooperativo em rede, visando ao controle dos metadados. Além disso, iniciamos a elaboração de uma Política de Indexação, que, dentre vários objetivos, um se refere ao controle dos metadados.

**8. Com que frequência se verifica a autenticidade e a integridade dos documentos?**

R: Não há uma verificação em relação à autenticidade ou integridade dos documentos após a entrada no sistema, porque apenas os administradores do sistema podem submeter novos dados/informações; portanto, não há risco de haver manipulação dos dados/informações armazenados nos sistemas.

**9. Quais são as estratégias adotadas para tal verificação?**

R: Não há verificação após a submissão inicial.

**10. Como a biblioteca demonstra a seus usuários a autenticidade e a integridade de um documento?**

R: Não há demonstração explícita.

**11. Já foi constatado algum tipo de alteração em um documento? Em caso positivo, como a alteração foi detectada? Qual foi o procedimento a partir dessa descoberta?**

R: Sim, uma administradora do sistema modificou o arquivo original de uma dissertação, a pedido da própria autora (que era funcionária da Rede de Bibliotecas). Isso ocorreu sem a ciência da Coordenação. Quando ficamos sabendo do ocorrido, elaboramos uma instrução de serviço que orienta os procedimentos de maneira que impeça que isso ocorra novamente.

**12. Qual é a periodicidade da avaliação das políticas de preservação? Quem são os responsáveis pela avaliação?**

R: Como não possuímos uma política de preservação formal, mas temos plena consciência de sua importância para a Rede de Bibliotecas, certamente desenvolveremos uma política que defina os responsáveis, as formas e a frequência de avaliação, mas por enquanto isso não ocorre.

**13. As políticas para preservação da autenticidade e da integridade já precisaram sofrer alguma alteração desde sua criação? Em caso positivo, por quê?**

R: Não possuímos uma política de preservação formal. Mas como já foi relatado anteriormente, houve a necessidade de modificar uma instrução de serviços, visando a manter a preservação e autenticidade de documentos eletrônicos/digitais.

**14. Como foram registradas essas alterações?**

R: Foi elaborada uma nova instrução de serviço.

**15. As políticas são divulgadas aos usuários? Em caso negativo, por quê? Há intenção de divulgá-las?**

R: Como não possuímos uma política formal, não há divulgação dela aos usuários. Contudo, todos os formulários e instruções normativas, relativas aos aspectos de preservação digital, são disponibilizados aos usuários da Rede de Bibliotecas.

**Se for possível, favor, fornecer uma cópia da atual política de preservação.**

R: Não possuímos uma política de preservação formalizada.

## **ANEXO D**

### **Respostas recebidas da Unicamp**

**1. Existe uma política especialmente criada para a preservação digital em longo prazo do acervo da biblioteca?**

**2. Estão determinadas na política as estratégias escolhidas para garantir a preservação do acervo?**

- Caso não exista uma política, ainda assim são adotadas estratégias de preservação? Quais?

R: Como a política de preservação digital encontra-se em fase de estudo e definições do modelo a ser adotado, e mesmo considerando o sistema de *backup* como não sendo um padrão correto de preservação digital, é o que está sendo realizado, da seguinte forma:

1. *Backup* incremental diário, dos novos itens inseridos no Banco Digital, dados e objeto digital.
2. *Backup* total do banco digital realizado nos fins de semana.

Todos os dados e objetos digitais são gravados em DVD com duas cópias, que são armazenados em locais fisicamente separados.

**3. Existe alguma explicação sobre as estratégias e/ou o porquê de sua escolha?**

R: Não, ainda não temos uma política definida e implantada para preservação digital. Existem estudos que estão sendo realizados por um grupo de trabalho da universidade, do qual a equipe da Biblioteca Digital participa, denominado “Projeto de Gestão, Acesso e Preservação de Documentos Arquivísticos Eletrônicos”. Este grupo de trabalho tem como objetivo elaborar normas que determinem padrões básicos de garantia à gestão, à preservação e ao acesso de documentos arquivísticos em meio eletrônico, gerados em cumprimento das funções da Unicamp. Essas normas serão utilizadas pelos órgãos e unidades para a produção e/ou a gestão de sistemas informatizados. Para dar continuidade aos estudos e às definições sobre a gestão de documentos digitais, tendo como um dos itens a preservação digital, a Unicamp estabeleceu convênio com o Arquivo Nacional para participar do Projeto InterPARES. Este convênio tem por objetivo: “Estabelecer parceria do Arquivo Nacional, órgão subordinado à Secretaria Executiva da Casa Civil da Presidência da República, com a Universidade Estadual de Campinas para fomentar o desenvolvimento de uma política de gestão e preservação de documentos arquivísticos digitais, com base no estudo de um conjunto de documentos arquivísticos digitais, a ser identificado pela Unicamp, para o qual serão propostas e testadas ações de gestão e preservação que garantam a autenticidade e a acessibilidade desses documentos, em conformidade com as recomendações do Projeto Internacional de Pesquisa sobre Autenticidade dos Documentos Digitais – InterPARES 3

(International Research on Permanent Authentic Records in Electronic Systems) de responsabilidade da Universidade de British Columbia. A identificação e a seleção do conjunto documental devem ter como critérios a relevância institucional para a tomada de decisão, o contexto organizacional, funcional, tecnológico e cultural”.

#### **4. Os conceitos de autenticidade e de integridade de documentos digitais estão presentes na política de preservação?**

- Em caso positivo, quais foram as definições adotadas pela biblioteca para esses dois conceitos? Quais são as estratégias e procedimentos adotados para preservação?
- Em caso negativo, por quê?

R: Apesar de os estudos estarem bem adiantados, principalmente após o início da participação da Unicamp no InterPARES, ainda não temos estabelecidos os conceitos finais sobre autenticidade e integridade de documentos digitais.

#### **5. Como a biblioteca verifica a autenticidade e a integridade de documento a ser incluído no acervo?**

R: Todo documento recebido para ser publicado na Biblioteca Digital deve ter sua origem em algum órgão oficial da universidade. Para as dissertações e teses, existe processo interno devidamente regulamentado, que envolve as secretarias de pós-graduação das unidades de ensino e pesquisa, geradoras das dissertações e teses; a Pró-Reitoria de Pós-Graduação, que regulamenta e homologa a dissertação ou tese; e a Biblioteca Central, que atua como depósito legal das teses homologadas e é responsável pela publicação das teses na Biblioteca Digital.

#### **6. Quais são os metadados de preservação adotados? Como eles são gerados?**

#### **7. Quais são as estratégias adotadas para controlar os metadados?**

R: Para o gerenciamento da Biblioteca Digital, utilizamos o *software* NOURAU, desenvolvido na Unicamp com tecnologia de *Software* Livre. A partir da definição da utilização do NOURAU, foram desenvolvidos alguns aplicativos visando a estabelecer a importação exportação de metadados das dissertações e teses. Com o auxílio da Biblioteca YAZ da Linguagem PHP (Módulo *teses2.php*), o sistema busca as informações dos metadados MARC oriundos do Catálogo Referencial – Base ACERVUS – e atribui o preenchimento das posições (Autor, Título, Palavras-Chaves, etc.) predefinidas nas páginas prefaciais do Aplicativo NOU-RAU. Durante o processo de migração das informações para

outros locais (Ex.: BD UNICAMP *versus* BDTD), ocorre a junção de duas aplicações, que são os Metadados Administrativos e os Estruturais, ou seja, o OAI coleta, processa e transfere as informações em XML, MARCXML);

- Na BD UNICAMP são utilizados (MARC, XML e OAI).

Categorias de Metadados: 1- Descritivos ou intelectual: descrevem e identificam os recursos de informação (Handle, PURL, DC, **MARC**, HTML Meta tags, vocabulários controlados). 2 Estruturais: facilitam a navegação e a apresentação dos recursos eletrônicos, linguagens para expressar metadados (SGML, **XML**, EAD, MOA2). 3- Administrativos: facilita o gerenciamento de longo e curto prazo e o processamento de coleções digitais (MOA2, CEDARS, **OAI**s “OAI e OAI DC”).

**8. Com que frequência se verifica a autenticidade e a integridade dos documentos?**

**9. Quais são as estratégias adotadas para tal verificação?**

R: Como descrito anteriormente, todo documento publicado na Biblioteca Digital, principalmente as dissertações ou teses, é originário e autêntico de um órgão oficial da universidade. Um documento digital só é publicado após ser verificada sua integridade física, ou seja, investiga-se se o documento está completo ou faltando páginas. No caso das dissertações e teses, a cópia digital deve ser idêntica à cópia impressa, e esta é recebida da Pró-Reitoria de Pós-Graduação após sua homologação.

**10. Como a biblioteca demonstra a seus usuários a autenticidade e a integridade de um documento?**

R: A Biblioteca Digital da Unicamp é um serviço oficial e devidamente regulamentada pela universidade. Ela foi criada pela Portaria do Gabinete do Reitor nº 85 de 8/11/2001. As atividades de publicação de dissertações e teses estão regulamentadas pelas normativas da Comissão Central de Pós-Graduação n<sup>os</sup> 001/2002, 003/2006, 051/2006 e pelo parecer da Procuradoria Geral da Universidade PG nº 3325/2002. Com base nessas regulamentações, os documentos publicados na Biblioteca Digital são autenticados de forma a garantir sua integridade.

**11. Já foi constatado algum tipo de alteração em um documento? Em caso positivo, como a alteração foi detectada? Qual foi o procedimento a partir dessa descoberta?**

R: Sim, e citamos alguns exemplos: documentos sem algumas páginas, o título do documento digital não coincide com o do impresso. A Biblioteca Digital da Unicamp procura seguir um



processo que garanta a qualidade do documento publicado. É certo que isso não representa garantia de 100%; quando ocorre alguma inconsistência de determinado documento publicado, busca-se o documento digital original e, ao se verificar que o original está alterado, tenta-se um contato com o seu autor, ou digitalizam-se as partes corretas do documento impresso para que sejam incorporadas ao digital. Após a criação de um novo arquivo digital correto, ele sobrescreve o que está armazenado, mantendo as informações de visitas e *downloads*. No momento da substituição do arquivo, fica registrada nos dados do documento a data da alteração realizada.

**12. Qual é a periodicidade da avaliação das políticas de preservação? Quem são os responsáveis pela avaliação?**

**13. As políticas para preservação da autenticidade e da integridade já precisaram sofrer alguma alteração desde sua criação? Em caso positivo, por quê?**

**14. Como foram registradas essas alterações?**

**15. As políticas são divulgadas aos usuários? Em caso negativo, por quê? Há intenção de divulgá-las?**

R: Como ainda não temos uma política oficial de preservação de documentos digitais, em razão de ela estar em fase de estudo e definição e vai servir a todos os órgãos da universidade, não temos como estabelecer periodicidade de avaliação, propor alterações, registrar possíveis alterações nem divulgar a todos os usuários. Destacamos que utilizamos uma política rígida de *backup* de todo o banco da Biblioteca Digital em diversos formatos e variações, não só do que foi incrementado no dia, como também do banco como um todo, diariamente. Esse sistema de segurança é utilizado não só para os dados dos documentos como também para o *software* gerenciador. Mantemos ainda em controle à parte, em um banco incremental, os dados de todos os usuários que realizaram *downloads* nas dissertações e teses, com os dados do cadastro realizado pelo usuário, a sua senha, qual(is) a(s) tese(s) foi(ram) copiada(s) e com a identificação da hora dessa atividade e do número do IP do equipamento que realizou o *download*. Esses dados de usuários são estritamente sigilosos e não são divulgados, somente são fornecidos se ocorrer alguma demanda da universidade.

#### **Documentos:**

Regulamentações da Biblioteca Digital: <http://www.sbu.unicamp.br/bibdig/regulamento.htm>

Instruções para Autores: <http://www.sbu.unicamp.br/bibdig/instrucao.htm>

Tese: Análise dos padrões e indicadores de qualidade para disponibilização das teses e dissertações na biblioteca digital da Unicamp.

<http://libdigi.unicamp.br/document/?code=vtls000339958>

Informações do InterPARES: <http://www.interpares.org/>

**Se for possível, favor, fornecer uma cópia da atual política de preservação.**