



Preservação digital em repositórios institucionais: práticas na região Sudeste do Brasil

Digital preservation in institutional repositories: practices in Southeastern Brazil

Simone da Rocha Weitzel*

Marco Aurélio Alencar de Mesquita**

RESUMO

Tem por objetivo a identificação das boas práticas de preservação digital aplicada aos repositórios institucionais (RIs), em instituições públicas de ensino e de pesquisa na região Sudeste do Brasil. Ressalta o papel da política de preservação digital nas práticas de preservação. Utiliza a pesquisa documental como método de coleta de dados nos RIs selecionados, e questionário de perguntas fechadas aplicados às instituições. Realiza análise quantitativa dos resultados, destacando as divergências aferidas entre os resultados das pesquisas. Estabelece, baseado na literatura e nos resultados obtidos, subsídios para o desenvolvimento de política de preservação digital.

Palavras-chave: Preservação Digital; Repositório Institucional; DSpace; Estratégias de Preservação Digital.

ABSTRACT

Aims to identify good practices of digital preservation applied to institutional repositories (IRs) in public institutions of education and research in the Southeast of Brazil. Highlights the benefits of digital preservation policy in the establishment of preservation practices. It is based on documental research methods such as data collection in selected IRs, and questionnaires in a closed questionnaire format, applied to institutions. Based on the technical literature and research results, it establishes a contribution towards the development of digital preservation policy.

Keywords: Digital Preservation; Institutional Repository; DSpace; Digital Preservation Strategies

INTRODUÇÃO

Os benefícios que a introdução das tecnologias da informação e comunicação (TIC) trouxeram para os sistemas de comunicação e produção científica são inegáveis e inovadores. Estes benefícios alcançam também os repositórios institucionais (RIs), que contribuem para a disseminação e democratização do acesso à informação. Contudo, a facilidade na divulgação de documentos proporcionada pela TIC por meio

* Professora associada da Universidade Federal do Estado do Rio de Janeiro. Doutora em Ciência da Informação. Endereço: Avenida Pasteur, 458, prédio do CCH, sala 404, CEP 22290-240, Rio de Janeiro, RJ. Telefone (21) 2542-1656. E-mail: sweitzel@unirio.br.

** Bacharel em Biblioteconomia pela Universidade Federal do Estado do Rio de Janeiro (Unirio) e Bibliotecário do Banco de Imagens Tyba Online. E-mail: marcoaum@gmail.com.

dos RIs pode esconder percalços próprios ao de um acervo, tal como o bibliográfico, no tocante à preservação e à garantia do acesso.

Conforme aponta Cunha (1999), a preservação material do papel requer espaço físico, condições climáticas, cuidados no manuseio, controle de empréstimo, além de custos com a biblioteca (manutenção do prédio, água, luz). Parece certo que com o novo paradigma digital – ou *hyper-paradigm*, como define Goicoechea (2007), em que o prefixo “*hyper*” representa os hipertextos –, a concentração do acervo em um RI aparenta, em um primeiro momento, a solução para diversos desses problemas. Ela representa, na verdade, a modificação de alguns aspectos de preservação que não podem – ou não devem – ser ignorados, como o estabelecimento de protocolos gerais de *backup* dos arquivos de sistema ou até mesmo de um módulo híbrido de trabalho, no caso de uma interrupção parcial do serviço.

O Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) investiu em várias iniciativas alinhadas ao Movimento do Acesso Aberto para apoiar a implantação de repositórios institucionais no Brasil nas últimas décadas – seja desenvolvendo a tradução do *software* livre DSpace e incentivando sua ampla adoção, seja com o lançamento de editais da Finep garantindo a distribuição de kits tecnológicos e apoio técnico-operacional para as instituições contempladas pelos editais na implantação de RIs.

Essas estratégias geraram frutos importantes para o país, tanto no sentido de potencialização da produção científica nacional, quanto em termos de participação mundial com conteúdos em português, especialmente de acesso aberto. No entanto, dados do OpenDOAR revelam que entre 2.509 repositórios ali cadastrados somente, 7,4% possuem uma política de preservação digital formalizada, ao passo que no Brasil há mais de 80% de RIs sem uma política de preservação definida e cerca de 10,8% sem política formalizada. Considerando que, de acordo com Sayão (2007), a preservação digital pode ser entendida como o conjunto de estratégias e protocolos na unidade da informação que visam à manutenção do banco de dados e dos objetos digitais disponibilizados, devidamente formalizados, não há como desenvolver ambientes com conteúdos digitais sem pensar em sua preservação. Este conjunto de estratégias e protocolos representa um compromisso de longo prazo com o acesso e a preservação (SAYÃO, 2007), sendo sua formalização imprescindível para a garantia da eficácia do sistema. Assim, o interesse em identificar quais seriam as principais ações de preservação que os RIs brasileiros vêm desenvolvendo configurou o problema de pesquisa.

Tendo em vista esse contexto, o presente estudo analisou as políticas de preservação digital em repositórios institucionais (RIs) situadas na região Sudeste brasileira que utilizam como base tecnológica o DSpace, por meio de pesquisa documental nos RIs selecionados e questionário de perguntas fechadas aplicados às instituições. Para estabelecer uma amostra das instituições a serem pesquisadas, partiu-se dos mesmos critérios de seleção de instituições elegíveis adotados para um edital de apoio a projetos de implantação de repositórios institucionais. Neste caso, foram utilizados os critérios do Edital Finep/PCAL/XBDB 003/2009, a saber:

- a) Ser uma instituição pública de ensino e pesquisa com atuação em território nacional.
- b) Manter programas de pós-graduação reconhecidos pelo MEC/Capes.
- c) Ter o compromisso com o modelo de interoperabilidade Open Archives.

Devido às restrições de cobertura do tema, a amostra limitou-se à região Sudeste.

A pesquisa teve como objetivo principal a identificação das boas práticas de preservação digital aplicada aos repositórios institucionais das instituições públicas de ensino e pesquisa na região Sudeste do Brasil, a fim de contribuir com subsídios para a elaboração de uma política específica para os RIs. Para isso, foi necessário:

- a) Identificar as estratégias de preservação digital usada – isoladamente ou combinadas – nos RIs das instituições públicas de ensino e de pesquisa.
- b) Identificar se a política de preservação digital está formalizada.
- c) Identificar como esta política está estabelecida: se é uma política da unidade de informação; se é da organização ou de parte de uma rede de colaboração.
- d) Identificar a interação dos RIs com iniciativas de preservação digital;

A partir dos resultados identificados, confirma-se a situação dos RIs brasileiros em relação à inexistência de política, uma vez que dentre os nove RIs pesquisados, somente um possui a política formalizada. Esse fato pode ser responsável por práticas pobres, tais como o baixo índice do uso de extensões de arquivos de alta adequação de preservação digital e a falta de diversificação nas estratégias de preservação digital, privilegiando: *backup*; inclusão de metadados; e preservação da tecnologia.

Espera-se que o presente estudo possa contribuir para futuras pesquisas, sobretudo para a análise do *status quo* dos RIs em relação à preservação digital, bem como para sensibilizar gestores de RIs da necessidade de engajamento nessa área, além de orientar na elaboração de políticas para fomentar as melhores práticas em preservação digital.

PRESERVAÇÃO DIGITAL EM REPOSITÓRIOS INSTITUCIONAIS

O conhecimento produzido, preservado e transmitido às gerações futuras pode ser considerado um dos grandes diferenciais da humanidade. O conhecimento preservado possibilitou o avanço da ciência, que se caracteriza por ser “uma instituição social, dinâmica, contínua, cumulativa” (TARGINO, 2000, p. 2).

Devido a essas características, a preservação e o acesso do conhecimento científico produzido ao longo de gerações de pesquisadores impulsiona a criação de um novo conhecimento para as gerações futuras. A ressignificação de um objeto na ciência moderna pressupõe avanço, confirmação ou refutação de uma pesquisa já realizada, conforme nos explica Abrantes e Martins:

A construção praxica [sic] do conhecimento nos remete, portanto, à realidade histórica a se conhecer, visto que os indivíduos se desenvolvem em relações de apropriação da história contida nos objetos produzidos pelo homem e nas relações estabelecidas entre eles na base de tais produções (ABRANTES; MARTINS, 2007, p. 315).

A necessidade de se salvaguardar o conhecimento registrado remonta às mais antigas civilizações. Vidal (2010, p. 146) destaca o fato de que “desde a antiguidade a Humanidade procurou, nos mais diversos suportes, guardar, conservar e preservar a sua memória”. Independentemente do suporte do registro da informação, a necessidade de preservá-los sempre foi um sentimento comum aos homens.

Apesar de a própria palavra biblioteca “designar um local onde se armazenam livros” (MORIGI; SOUTO, 2005, p. 190) e, ainda povoar, no imaginário popular, o papel da biblioteca clássica, definida como “aquela onde a maioria dos itens do seu acervo é constituída de documentos em papel” (CUNHA, 1999, p. 258), pouco se imagina que as bibliotecas sempre atuaram na vanguarda da tecnologia, pois:

A biblioteca possui, como instituição social, uma longa e complexa história pouco conhecida pela maioria das pessoas. As tecnologias da imprensa, máquina de escrever, telefone, telex, mimeógrafo, microfilme, cartão perfurado nas margens, computador, disco ótico e redes eletrônicas afetaram e alteraram a biblioteca ao longo do tempo. Algumas dessas tecnologias, tais como o microfilme e o disco ótico, tiveram suas primeiras aplicações testadas dentro de uma biblioteca (CUNHA, 1999, p. 257).

Atualmente, conforme aponta Sayão (2005), grande parte dos documentos gerados já nasce digitalmente, mesmo não havendo “estratégias consolidadas para garantir o acesso de longo prazo às informações digitais” (SAYÃO, 2005, p. 116). Isso se dá pela facilidade em produzir – o que Sayão chama de compulsão – e compartilhar os documentos, que é infinitamente superior à capacidade de preservá-los. A sociedade contemporânea traz em si a urgência do acesso, da produção e do compartilhamento. Talvez seja esta uma hipótese para explicar porque os documentos atuais nascem digitalmente, e os que não nasceram digitalmente são convertidos ao novo formato.

Uma série de políticas que visam à modificação da relação entre os produtores da informação e os usuários vem sendo discutida, entre elas o Copyleft e o Movimento de Acesso Aberto. Este último é responsável pela criação de serviços, protocolos e políticas que visam, entre outros objetivos, à interoperabilidade dos provedores de dados e serviços. Um desses sistemas são os Repositórios Institucionais que

Inseridos no coração do movimento mundial em favor do acesso aberto à informação científica, [...] constituem de fato inovação no sistema de comunicação da ciência e no modo como a informação – aquela que alimenta e resulta das atividades acadêmicas e científicas – é gerenciada. (LEITE et al., 2012, p. 7).

Os repositórios institucionais são, segundo Santos Júnior (2010, p. 36), “coleções digitais que armazenam, preservam, divulgam e dão acesso à produção intelectual de uma determinada comunidade, onde os documentos são disponibilizados irrestritamente na *web*”. Ou seja, destaca-se entre as formas atuais de dinamização e democratização do acesso à informação, por suas características genéricas – um tipo de coleção que conta com diversos tipos de documentos e formatos divulgados na *web*, além de ser gerenciado por uma instituição que garante verificação e tratamento da informação prestada.

Os repositórios nascem da necessidade da criação de um espaço único de acesso às informações produzidas sobre uma determinada instituição, portanto multidisciplinar, sendo um ponto confiável de coleta de dados e informações, num ambiente controlado, hierarquizado e ordenado, associando as vantagens dos acervos digitais, como acesso aberto, por exemplo. Além disso, ele segue a lógica da Iniciativa dos Arquivos Abertos, em que os resultados de pesquisa produzidos com financiamento público por intermédio de agências de fomento, por exemplo, devem ser disponibilizados gratuitamente na *web*, caracterizando o que se convencionou denominar de Acesso Aberto Verde, por meio dos RIs (SANTOS JÚNIOR, 2010).

A PRESERVAÇÃO DIGITAL

A preservação é um fator primordial, tanto para os documentos impressos, entre outros materiais, quanto para os digitais. A garantia de acesso futuro não pode ser considerada um item inerente à automatização dos processos, mas, assim como nos documentos em papel, deve surgir a partir da “adoção de ferramentas que protejam e garantam a sua manutenção” (MÁRDERO ARELLANO, 2004, p. 17).

A preservação digital é um esforço que “exige políticas, estratégias e ações para garantir, no decorrer dos anos, acesso a conteúdo válido apesar dos desafios impostos pelos suportes (mídias) e mudanças tecnológicas” (CORRÊA, 2010, p. 23). Não é, portanto, fruto de uma ação isolada, mas de um esforço institucional para garantir o acesso futuro aos documentos digitais. Estas políticas devem existir, preferencialmente, por escrito, pois independentemente de serem utilizadas de modo individual, entre instituições que trabalham por cooperação, elas

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 o processo de preservação de acordo com padrões e melhores práticas que garantam uma custódia responsável (CORRÊA, 2010, p. 23-24).

Ao contrário dos documentos que apresentam suportes materiais, como o papel ou pergaminho, por exemplo, que em condições específicas podem ficar anos, décadas, séculos preservados, prontos para transferir a informação ali depositada à próxima pessoa que o acessar, os documentos digitais necessitam de vários elementos externos para realizar esta mesma transferência. Sendo assim, “para muitos documentos analógicos, a preservação do suporte e sua forma física garantem a conservação de seu conteúdo” (CORRÊA, 2010, p. 24), o que não pode ser garantido com os objetos digitais. Um livro pode ficar décadas sobre uma mesa – em condições específicas – e ser relido nos dias futuros, quando dificilmente uma música em uma fita K7 será ouvida hoje em dia, em lares onde já existem rádios com entradas USB. A preservação digital deve ter como principal meta garantir recuperação e visibilidade do documento digital no futuro (SANTHANAGOPALAN; FOX; McMILLAN, 2006, p. 1).

Isso tende a não acontecer “devido ao fato de que a informação tem um fator de dependência do seu ambiente técnico” (SANTHANAGOPALAN; FOX; McMILLAN, 2006, p. 1). Garantir a existência de um arquivo/formato pode não significar automaticamente a garantia de acesso aos arquivos nas mesmas condições em que foram produzidos, pois o *software* reprodutor daquele conteúdo pode ter sido atualizado, imputando modificações consideráveis ao arquivo. Essa obsolescência natural ou provocada (GRANGER, 2002) traz uma grande fragilidade aos documentos nascidos digitalmente.

ESTRATÉGIAS E INICIATIVAS EM PRESERVAÇÃO DIGITAL

As diversas estratégias de preservação digital (PD) existentes visam contribuir para a preservação do objeto digital e garantir o acesso futuro. As estratégias não comportam todos os itens previstos para uma PD, estão sempre focadas em um item específico do tema, evidenciando a complexidade do problema da preservação digital. Parece não haver uma estratégia única capaz dar conta de todos os processos

de preservação digital (SAYÃO, 2005), mas ainda assim é possível salientar pontos comuns para a preservação digital.

Na literatura, é possível identificar diversas estratégias de preservação digital, tais como: preservação da tecnologia, emulação, migração, encapsulamento, metadados (descritivos, estruturais e administrativos), identificadores persistentes, *backup*, redes de distribuição de preservação digital, por exemplo. Também se deve levar em consideração os tipos e extensões que povoarão o RI, para que seja minimizada a necessidade de migrações em massa.

As iniciativas em PD são ações conjuntas que visam produzir uma maneira racional e colaborativa de preservação digital. O esforço pela preservação digital envolve diversos fatores, sendo possível que fatores financeiros possam ser sinalizados como obstáculos na sua execução. As iniciativas de preservação digital surgem como um auxílio na tarefa extremamente trabalhosa e custosa para apenas uma instituição desenvolver.

Dentre estas iniciativas, podemos citar o LOCKSS (Lots of Copies Keeps Stuff Safe – em inglês, “Muitas cópias mantêm os materiais a salvo”, em tradução livre) que é um *software* de código aberto que permite a criação de redes privadas locais, em tradução livre para *Private LOCKSS Network* em inglês (PLN). As PLNs “preservam cópias de conteúdos digitais criados localmente por organizações científicas e culturais” (MÁRDERO ARELLANO, 2012, p. 85).

A Rede Cariniana é uma iniciativa do Ibict, que tem o “objetivo de propor a construção e customização de uma rede nacional de serviços de preservação digital” (MÁRDERO ARELLANO, 2012, p. 87), contando com a parceria de instituições de ensino e pesquisa e com a colaboração de especialistas brasileiros, nesta construção. Sua atuação reforça a necessidade da criação de iniciativas de preservação que garantam a salvaguarda dos registros digitais em âmbito nacional. Trata-se de uma importante iniciativa que busca orientar as instituições na prática da preservação digital de publicações de acesso aberto, oferecendo “pacotes de *softwares*, aplicações e ambientes multimídia para a implementação e desenvolvimento de documentos digitais preserváveis” (MÁRDERO ARELLANO, 2012, p. 87).

A operacionalização da rede é baseada no LOCKSS, adotando o modelo de DDP com pretensões de atingir redes de âmbito nacional. Proporciona um espaço para a busca de melhores metodologias e práticas em preservação digital, de forma colaborativa.

Estudos de casos

Para identificar as melhores práticas em preservação digital (PD) nos repositórios Institucionais (RIs) das instituições públicas de ensino e de pesquisa na região Sudeste, verificando de que maneira estas ações estão formalizadas como política, empregaremos dois métodos combinados para descrever os estudos de caso: a) pesquisa documental, para a coleta de dados nos RIs sobre as informações básicas referente às políticas e aos demais aspectos gerais da estratégia de PD; e b) questionário com questões fechadas para coletar dados mais específicos. Posteriormente, foi realizada uma análise quantitativa das respostas para identificar as boas práticas em relação à PD em RIs.

Os RIs foram selecionados para a amostra da pesquisa baseados nos critérios de instituições elegíveis do Edital Finep/PCAL/XBDB 003/2009, presentes na região Sudeste do Brasil. Estes critérios delimitam um escopo do perfil das instituições

selecionadas pela amostra, tendo em vista a diversidade de assuntos e atividades dos repositórios institucionais existentes, e garantindo a homogeneidade das instituições selecionadas e o uso de uma ferramenta comum: o DSpace, atendendo ao quesito da utilização do modelo de interoperabilidade Open Archives.

Os nove repositórios selecionados para a amostra estão assim distribuídos:

Quadro 1 – Amostra de Repositórios Institucionais.

tec	UF	Tamanho da Coleção
1	SP	43.350
2	MG	2.588
3	SP	338
4	SP	1.374
5	RJ	425
6	MG	4.237
7	SP	126
8	SP	231
9	RJ	225

Fonte: Elaborado pelos autores.

A pesquisa documental baseou-se na visita ao endereço eletrônico dos nove repositórios institucionais selecionados e listados acima para avaliar a disponibilidade de recursos e informações básicas aos usuários dos sistemas. Foram selecionados cinco perguntas baseados nos critérios do edital de fomento aos RIs e na literatura técnica referente ao assunto, com o objetivo de identificar a coleção, as políticas e estratégias de preservação digital, a saber: tamanho da coleção; política de preservação digital acessível; relação dos formatos de arquivos aceitos; identificação da participação em redes de preservação distribuída e se apresenta identificação persistente no *link* de endereço aos objetos digitais.

Foi desenvolvido um questionário com 12 questões fechadas, elaboradas com o objetivo de identificar as práticas das instituições com relação à preservação digital. Suas questões foram agrupadas em três blocos: identificação da instituição, política de preservação digital e estratégias de preservação digital.

O questionário foi desenvolvido por meio de um formulário *online*, utilizou recursos *online* – Google.docs – para a tabulação dos resultados e a coleta de dados, permitindo o acompanhamento das respostas em tempo real e ficando disponível para resposta durante um mês.

Apesar da receptividade dos bibliotecários em relação ao questionário, em alguns casos a resposta foi enviada após troca de diversas mensagens sobre a necessidade do levantamento para a conclusão desta pesquisa.

TABULAÇÃO E ANÁLISE DOS DADOS

Os resultados apresentados correspondem a uma amostra total de nove repositórios da região Sudeste, de instituições públicas de ensino e de pesquisa com atuação em território nacional, mantenedoras de programas de pós-graduação reconhecidos pelo MEC/Capes, conforme o critério de instituições elegíveis do Edital Finep/PCAL/XBDB 003/2009. Os resultados serão apresentados em duas seções: pesquisa documental e questionário.

O primeiro ponto a ser avaliado na pesquisa documental diz respeito ao tamanho da coleção. Foi possível identificar um grande número – cinco a nove RIs – que apresentam uma coleção menor que 500 registros, divergindo do resultado aferido no questionário, respondido pelos gestores dos RIs, conforme veremos mais à frente. Os resultados foram obtidos por meio do somatório dos valores encontrados nos endereços eletrônicos dos repositórios selecionados na amostra. Cada comunidade inserida no repositório traz o total de registros cadastrados em sua categoria, formando assim o número total de registros daquele repositório.

O segundo item avaliado na pesquisa documental foi a publicidade da política de preservação digital. Dos nove RIs visitados, oito não tornaram acessíveis suas políticas de preservação digital. Apenas um dos repositórios pesquisados torna a sua política disponível, mas em forma de resolução, não se aprofundando em questões técnicas. A mesma instituição, em resposta ao questionário, informou que há uma política de PD válida para toda a instituição, porém não tivemos acesso, por ser um documento interno. Não é possível descartar a possibilidade que essa política de fato exista nos demais RIs pesquisados, mas não por escrito.

Outro item analisado corresponde à indicação aos formatos autorizados para inclusão no RI. Não foi possível identificar em nenhum dos RIs pesquisados quais são, de fato, os formatos de arquivos disponíveis e passíveis de povoamento nos repositórios, sendo possível mensurá-los apenas no questionário.

Também por meio da pesquisa documental, foi possível identificar que não há qualquer participação dos RIs em redes de preservação distribuída. Mesmo que dois dos nove RIs pesquisados tenham declarado realizar este modo de preservação digital durante a resposta do questionário, como veremos a seguir não é possível identificar de maneira clara esta parceria no endereço eletrônico dos RIs pesquisados.

O último item analisado na pesquisa documental foi à apresentação de *link* persistente nos endereços dos objetos digitais, onde foi possível identificar um uso de 100% do Handle System.

Os dados apresentados a seguir foram coletados através de questionário de perguntas fechadas com nove RIs. Dos quatro estados da federação que compõe a região Sudeste do Brasil, apenas o Espírito Santo não possui RI contemplado na amostra. Então, a distribuição geográfica dos RIs pesquisados é dividida entre: duas unidades no Rio de Janeiro, duas em Minas Gerais e cinco no estado de São Paulo, conforme foi mostrado no Quadro 1.

Há uma concentração de RIs no estado do São Paulo. Este percentual é composto pelo maior número de unidades que correspondem aos critérios de seleção da pesquisa, distribuída também pelo interior do estado, ao contrário aos demais, que se concentram apenas na capital ou cidades-chave.

O questionário revelou que cinco dos nove RIs pesquisados possui um acervo superior à 2.000 registros, sendo um total divergente com o resultado aferido na pesquisa documental.

A amostra revelou também que oito dos nove RIs pesquisados não possuem uma política de preservação digital formalizada. A falta deste documento, formalizado – escrito de fato – pode evidenciar problemas em relação à padronização de critérios, métodos ou periodicidade definida para o desenvolvimento do processo de preservação digital. O conhecimento tácito não deve ser a única fonte de conhecimento para ações de política de preservação digital, conforme foi evidenciado pelos resultados obtidos. Somente um dos nove RIs possui uma política de preservação digital institucionalizada. A política de preservação digital, quando formalizada e pública, ajuda a sedimentar um posicionamento transparente do repositório em relação à preservação digital, orientando os autores (depositantes) a respeito dos procedimentos adotados e reforçando a garantia de perpetuação e acesso.

A participação em iniciativas em preservação digital em rede demonstrou um percentual negativo em sete dos nove RIs pesquisados. Dos dois que responderam positivamente, apenas a iniciativa da Rede Cariniana foi citada, sendo uma delas mencionada como “a do Ibict”, mesmo constando a Rede Cariniana entre as opções de resposta.

Em relação às estratégias de preservação digital utilizadas, um número significativo de RIs – sete – declararam utilizar o *backup* como estratégia de preservação digital, seguido de quatro menções à inclusão de metadados, sendo que apenas uma unidade declarou realizar as duas estratégias ao mesmo tempo, e um RI declarou realizar as duas anteriores e, ainda, a preservação da tecnologia. Assim, é possível identificar que as estratégias de *backup* e inclusão de metadados estão mais difundidas nos RIs, sendo citadas apenas três estratégias nos resultados. Não receberam nenhuma citação estratégias como migração, encapsulamento ou preservação digital distribuída.

A identificação persistente também revelou um resultado divergente ao apresentado na pesquisa documental. Houve a declaração de uso preferencial de dois sistemas de identificação: o Handle System – três RIs – e o DOI – dois RIs –, com os demais declarando não utilizar nenhum sistema de identificação persistente, o que representa uma considerável divergência com os resultados da pesquisa documental, que apresentaremos a seguir.

Foi levantada uma extensa relação de formatos de texto, imagem, áudio e vídeo, submetendo as extensões no questionário, para que os gestores dos RIs identificassem quais delas são aceitas no povoamento do RI.

Para a apresentação dos resultados dos arquivos utilizados, precisou-se de um arranjo diferenciado, devido à grande opção de respostas. Para isso, foram selecionados apenas as extensões que receberam alguma menção na resposta do questionário. Em seguida, os resultados foram separados em quadros onde aparecem subdivididos em três, a saber: alta, média e baixa adequação para preservação, classificação baseada na tabela de adequação para a preservação de Munhoz (2011).

Sendo assim, os resultados obtidos correspondem aos quadros a seguir:

Quadro 2 – Resultados – Tipos de Arquivos – Texto (todos os RIs).

	Extensão	Uso (Nº do RI)
Alta	PDF/A	1, 3, 4, 6, 8 e 9
	TXT	2, 3, 6 e 9

Alta	HTML, XHTML ou XML	2 e 3
	ODT	3
Média	PDF	Todos
	DOCX	3, 4, 6 e 9
	RTF	6 e 9
	CSS	3
Baixa	DOC	3, 4, 6 e 9
	Todos os formatos	4

Fonte: elaborado pelos autores.

É possível identificar uma forte tendência na utilização dos arquivos em PDF/A e PDF, mesmo sendo de classificação de adequação para preservação diferentes, alta e média, respectivamente.

A análise de extensões de imagens levou em consideração os cinco RIs que não aceitam imagens (os RIs 1, 5, 7, 8 e 9), não classificando seu resultado em nenhuma categoria. Também se levou em consideração os RIs que declararam aceitar qualquer tipo de imagem, classificando esse resultado como de baixa adequação para a preservação digital, por ser uma prática nociva ao próprio RI.

Quadro 3 – Resultados – Tipos de arquivos – Imagem.

	Extensão	Uso (Nº do RI)
Alta	TIF	2 e 3
	PNG	3
Média	JPG	2, 3, 4, 6
	GIF	3 e 6
	BMP	3
Baixa	PSD	6
	TIFF	3
	Todos os formatos	4

Fonte: elaborado pelos autores.

Assim, é possível identificar uma disparidade entre o uso das extensões com alta e média adequação. Além disso, das quatro opções de extensão disponíveis para alta adequação, conforme Munhoz (2011), apenas duas foram citadas.

A análise dos resultados para as extensões de vídeo não considerou os RIs que não aceitam este tipo de mídia (RIs 5, 7, 8 e 9). Os resultados demonstram uma maior tendência no uso de arquivos de média adequação, apesar do também elevado índice de alta adequação.

Quadro 4 – Resultados – Tipos de arquivo – Vídeo.

	Extensão	Uso (Nº do RI)
Alta	MOV	2 e 3
	AVI	2 e 4
Média	MPEG	2, 3, 4 e 6
	MP4	1, 2, 4
Baixa	Todos os formatos	4

Fonte: elaborado pelos autores.

O resultado também demonstra que apenas quatro dos dez formatos disponíveis (sem levar em consideração as variações), conforme Munhoz (2011), foram citados nas respostas, o que sinaliza a baixa adesão deste formato nos RIs. Grande parte das extensões de vídeos declaradas em resposta ao questionário foi de média adequação.

Desse modo, os formatos citados durante o questionário correspondem ao quadro a seguir apresentado, culminando na amostra consolidada dos formatos utilizados nas instituições públicas de ensino e pesquisa com atuação em território nacional, mantenedora de programas de pós-graduação reconhecidos pelo MEC/Capes, conforme critério de edital de fomento aos RIs.

Quadro 5 – Resultado Consolidado.

	Texto	Imagem	Uso (Nº do RI)
Alta	HTML, XHTML ou XML TXT PDF/A ODT	PNG TIF	MOV AVI
Média	PDF CSS RTF DOCX	GIF JPG BMP	MPEG MP4
Baixa	DOC	PSD TIFF	---

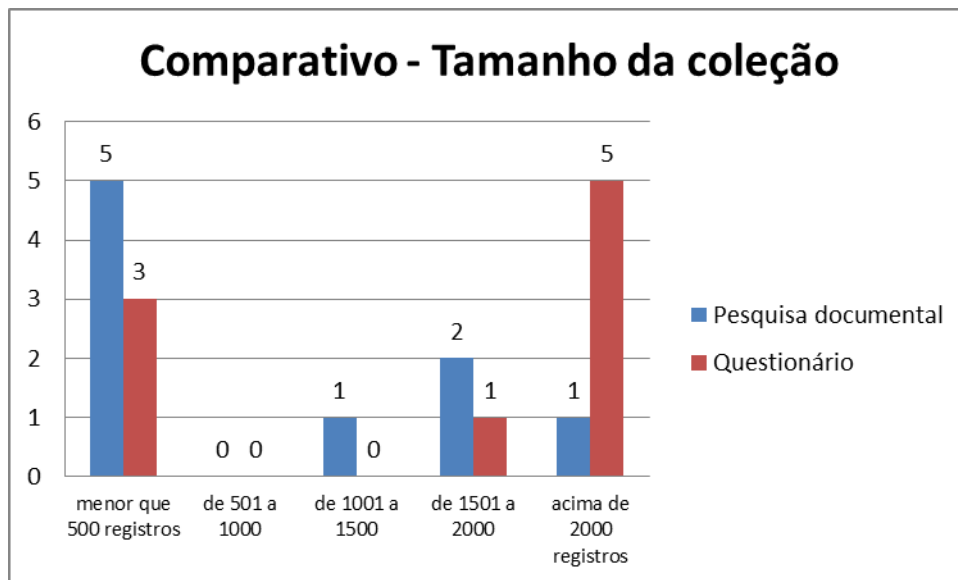
Fonte: elaborado pelos autores.

Entre os resultados obtidos durante o estudo de caso, é importante ressaltar a discrepância em dois: tamanho da coleção e identificador persistente.

Por intermédio da pesquisa documental e dos resultados do questionário, é possível identificar uma polarização inversa entre os resultados do quesito “tamanho da coleção”. Enquanto na pesquisa documental é revelado o maior índice de

repositórios com uma coleção menor que 500 registros, os resultados do questionário revelam o maior índice de repositórios com uma coleção acima de 2.000 registros, conforme gráfico abaixo.

Gráfico – Resultados – Tamanho da coleção.



Fonte: elaborado pelos autores.

Por meio do gráfico comparativo, é possível identificar uma acentuada divergência entre o que os gestores declaram como sendo o acervo integral que povoa os RIs e o total que está de fato disponível para consulta

A identificação persistente também revelou um resultado divergente entre a pesquisa documental e o questionário. Confrontando os resultados obtidos na resposta ao questionário – em que o Handle System obteve três menções, e o DOI, duas, ao passo que os outros declararam não utilizar nenhum tipo de identificação persistente – com os resultados da pesquisa documental, verificou-se uma grande diferença. Durante a pesquisa documental, foi possível identificar 100% de uso do Handle System como método de identificação persistente dos objetos digitais, pelo uso do endereço “.../handle/” nas URLs dos que foram depositados nos RIs selecionados pela amostra.

BOAS PRÁTICAS DE PRESERVAÇÃO DIGITAL

Este trabalho teve por objetivo identificar as boas práticas de preservação digital. Baseando-se nos resultados obtidos e no conjunto de objetivos genéricos apresentados por Bullock (1999), define-se que a política de preservação digital prescindiria de três elementos básicos para sua esquematização: o objeto digital, os metadados e a política.

O objeto digital que faz parte de um RI precisa ter a garantia de sua preservação em longo prazo, sem perda de conteúdo ou modificação na sua apresentação. Para tanto, é necessário que se estabeleçam critérios durante o processo de povoamento desses objetos, com a escolha de formatos que possibilitem um bom grau de adequação às práticas de preservação.

Os metadados devem ter uma participação ainda maior no processo de preservação digital, ampliando seu uso para além dos metadados descritivos. É necessário o uso de metadados estruturais e administrativos, complementando a função descritiva do objeto digital no contexto do RI.

A política de preservação formalizada deve ser o instrumento para orientar as práticas de preservação digital, sendo institucionalizadas ou não. Uma vez formalizada, ela explicitará para a comunidade atendida pelo RI os procedimentos de preservação, garantindo sua sistematização e continuidade em longo prazo. Poderá, ainda, garantir a autenticidade e a proveniência de termos de publicação e outros instrumentos que podem ser associados a ela.

Dessa forma, a presente pesquisa identificou os seguintes critérios que devem ser considerados nas propostas de políticas de preservação digital:

- Escolha de formatos de alta e média adequação (visando minimizar a necessidade de migração).
- Uso de estratégias de preservação como, por exemplo, a inclusão de metadados descritivos (como título, autor e palavras-chave), de backup e, se possível, de estratégias de formas combinadas.
- Uso de metadados estruturais (auxiliam-nos na estruturação dos documentos digitais, preservando a sua apresentação) e administrativos (que auxiliam na gestão do ciclo de vida dos objetos digitais, possibilitando a inclusão de informações de criação, especificidades e dependências técnicas).
- Uso de identificação persistente.
- Adesão a uma rede distribuída para preservação digital.
- Uso de uma política de preservação formalizada.

Esses elementos poderão ser utilizados como subsídios para a elaboração de uma política de preservação digital.

CONSIDERAÇÕES FINAIS

A pesquisa realizada buscou ressaltar aspectos importantes das estratégias de preservação digital utilizadas pelos RIs, com o objetivo de mapear, da maneira mais fiel possível, as práticas de preservação digital utilizadas no dia a dia, evidenciando possíveis contribuições para a formalização de um subsídio para políticas de preservação em repositórios institucionais.

Destacam-se a colaboração de todos os bibliotecários responsáveis pelos RIs na construção desta pesquisa e no esforço contínuo para, cada vez mais, possibilitar o acesso à informação de forma ágil, completa e irrestrita.

Os resultados obtidos por meio da pesquisa documental e do questionário demonstraram ser suficientes na identificação dos objetivos da pesquisa, ajudando a delimitar as estratégias de preservação digital das unidades selecionadas.

Foi possível também identificar que oito dos nove RIs pesquisados não possuem uma política de preservação digital formalizada, e apenas um RI pesquisado apresenta uma política institucionalizada. Esse percentual atesta o uso inexpressivo de um recurso tão importante como a política de preservação digital, estando aquém dos resultados desejados e demonstrando o distanciamento dessa política dos

demais objetivos da universidade/instituição para qual ela presta serviço, e que, certamente, tem interesse na preservação do conhecimento produzido. O baixo índice no uso de extensões de arquivos de alta adequação de preservação digital pode refletir a falta de políticas na escolha desses arquivos.

A política de preservação digital é um importante instrumento para a garantia do cumprimento da estratégia de preservação digital. Porém, ficou evidente que este instrumento não está presente em grande parte dos RIs pesquisados. É possível presumir que haja atividades de preservação digital nas unidades pesquisadas; contudo, não foi possível mensurar sob quais circunstâncias ela se dá ou se realmente as ações são cumpridas.

Apurou-se um alto percentual de RIs que não participam de redes de preservação digital: sete de nove RIs. A política de preservação em redes oferece enormes benefícios, tais como a garantia perene de acesso aos objetos digitais, uma vez que eles se encontrariam salvaguardados em mais de uma unidade participante da rede.

Também ficou evidente nos resultados da amostra a falta de diversificação nas estratégias de preservação digital. Apenas três estratégias são citadas: *backup*, inclusão de metadados e preservação da tecnologia. Além disso, apenas dois RIs pesquisados declararam utilizar mais de uma estratégia como medidas complementares.

Considerando o baixo número de políticas institucionais formalizadas, a baixa adesão às estratégias de preservação em rede e a falta de diversificação das estratégias de preservação digital, os resultados apresentados comprovam a fragilidade das coleções digitais no que se refere à preservação digital.

É necessário um esforço em nível nacional nas instituições que mantêm algum RI, a fim de que se estabeleçam diretrizes sólidas para os critérios de criação de uma política de preservação digital.

Cabe ressaltar que mesmo no Edital Finep/PCAL/XBDB 003/2009, usado como parâmetro no critério para a escolha de instituições elegíveis para a amostra desta pesquisa, o tema da preservação digital não é abordado – ou mesmo uma garantia de manutenção mínima do conteúdo –, ainda que o objetivo do edital seja selecionar projetos de apoio à implantação de RIs. É certo que o mesmo edital estabelece compromissos – tais como a conexão permanente na internet e a criação de uma política institucional de informação que vise garantir a alimentação do RI por parte dos pesquisadores das instituições –, mas estas medidas garantem, respectivamente, apenas o acesso e o povoamento, e não a preservação digital. Por consequência, elas não se constituem em garantia da manutenção do conhecimento produzido para as gerações futuras.

A necessidade em produzir – por parte dos autores – e de organizar e promover acesso a essa produção – por partes dos RIs – deve encontrar o equilíbrio com a preservação digital, para garantir que o esforço dispensado hoje esteja acessível no futuro.

Os recursos eletrônicos representam, de fato, um avanço na comunicação científica, possibilitando acesso rápido e contínuo aos acervos dos RIs. Entretanto, é necessário lembrar que os recursos tecnológicos não estão imunes a problemas, tanto no que se refere aos *softwares* quanto aos *hardwares*. Estes problemas podem comprometer o acesso parcial ou total ao acervo, limitando o seu uso. É necessária a criação de estratégias que visem à garantia da permanência de acesso aos objetos digitais, do modo como foram desenvolvidos, e, também, aos registros dos bancos de dados dos

RIs, para que, havendo qualquer falha, o restabelecimento possa se dar da forma mais rápida possível.

As estratégias apresentadas neste trabalho não correspondem, necessariamente, à necessidade de um elevado investimento financeiro para que a preservação digital seja realizada a contento, mas elas dependem de políticas formalizadas e sedimentadas nas práticas das instituições que detenham um RI. O homem atribui ao computador a responsabilidade pela preservação, como se a migração para o suporte digital já representasse este fim. É necessária a criação de espaços de discussão e troca de experiências para reforçar a importância das práticas de preservação digital.

Artigo recebido em 31/01/2015 e aprovado em 10/04/2015.

REFERÊNCIAS

ABRANTES, A. A; MARTINS, L.M. A produção do conhecimento científico: relação sujeito-objeto e desenvolvimento do pensamento. *Interface: comunicação, saúde e educação*, Botucatu, SP, v. 11, n. 22, p. 313-325, maio/ago. 2007.

BULLOCK, A. Preservation of digital information: issues and current status. *Network Notes*, Ottawa, n. 60, abr. 1999. Disponível em: <<http://epe.lac-bac.gc.ca/100/202/301/netnotes/netnotes-h/notes60.htm>>. Acesso em: 1º maio 2014.

CORRÊA, A.M.G. *Preservação digital: autenticidade e integridade de documentos em bibliotecas digitais de teses e dissertações*. 2010. 96 f. Dissertação (Mestrado em Ciência da Informação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2010. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/27/27151/tde-05112010-105831/publico/2916162.pdf>>. Acesso em: 12 abr. 2014.

CUNHA, M.B. Desafios na construção de uma biblioteca digital. *Revista Ciência da Informação*, Brasília, v. 28, n. 3, p. 257-268, set./dez. 1999.

GOICOECHEA, M. Hyper-paradigm. In: SANZ, A.; ROMERO, D. (Edit.). *Literatures in the digital age: theory and praxis*. Newcastle, UK: Cambridge Scholars Publishing, 2007. p. 18-21.

GRANGER, S. Digital preservation and deep infrastructure. *D-Lib Magazine*, v. 8, n. 2, fev. 2002. Disponível em: <<http://www.dlib.org/dlib/february02/granger/02granger.html>>. Acesso em: 10 fev. 2014.

LEITE, F. et al. *Boas práticas para a construção de repositórios institucionais da produção científica*. Brasília: Ibict. 2012. p. 34. Disponível em: <<http://livroaberto.ibict.br/handle/1/703>>. Acesso em: 24 abr. 2014.

MÁRDERO ARELLANO, M. A. Cariniana: uma rede nacional de preservação digital. *Revista Ciência da Informação*, Brasília, v. 41, n. 1, p. 83-91. jan./abr. 2012. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/2127>>. Acesso em: 23 jul. 2014.

MÁRDERO ARELLANO, M. A. Preservação de documentos digitais. *Revista Ciência da Informação*, Brasília, v. 33, n. 2, p. 15-27. maio/ago. 2004.

MORIGI, V.J; SOUTO, L.R. Entre o passado e o presente: as visões de biblioteca no mundo contemporâneo. *Revista ACB: biblioteconomia em Santa Catarina*,

Florianópolis, v. 10, n. 2, p. 189-206. jan./dez. 2005. Disponível em: <<http://revistaacb.emnuvens.com.br/racb/article/view/432>>. Acesso em: 26 abr. 2014.

MUNHOZ, A. M. *Preservação digital de acervos: uma análise das políticas e estratégias utilizadas pelas universidades do Rio Grande do Sul*. 2011. 63 f. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia) – Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul. Porto Alegre, 2011. Disponível em: <<http://hdl.handle.net/10183/37473>>. Acesso em: 10 set. 2014.

SANTHANAGOPALAN, K; FOX, E.A; McMILLAN, G. A prototype for preservation and harvesting of international ETDs using LOCKSS and OAI-PMH. In: INTERNATIONAL SYMPOSIUM ON ELECTRONIC THESES AND DISSERTATIONS, 9., 2006. ,Québec, Canadá. *Anais eletrônicos...* Quebec City: NDLTD, 2006. Disponível em: <http://scholar.lib.vt.edu/theses/presentations/ETD_2006_Kamini_Santhanagopalan.doc>. Acesso em: 12 abr. 2014.

SANTOS JUNIOR, E. R. *Repositórios institucionais de acesso livre no Brasil: estudos delfos*. 2010. 177 f. Dissertação (Mestrado em Ciência da Informação) – Universidade de Brasília. Brasília, 2010. Disponível em: <http://repositorio.unb.br/bitstream/10482/5343/6/2010_ErnaniRufinodosSantosJunior.pdf>. Acesso em: 12 abr. 2014.

SAYÃO, L.F. Preservação digital no contexto das bibliotecas digitais: uma breve introdução. In: MARCONDES, Carlos Henrique et al. (Org.). *Bibliotecas digitais: saberes e práticas*. Salvador: Ufba; Brasília: Ibict, 2005. p. 115-146.

SAYÃO, L.F. Conservação de documentos eletrônicos. In: GRANATO, M.; SANTOS, C. P.; ROCHA, C. R. A. (Org.). *Conservação de acervos*. Rio de Janeiro: Mast, 2007. p. 181-204. (Mast Colloquia; 9). Disponível em: <http://www.mast.br/livros/mast_colloquia_9.pdf>. Acesso em: 10 fev. 2014.

TARGINO, M.G. Comunicação científica: uma revisão de seus elementos básicos. *Revista Informação & Sociedade*, João Pessoa, v. 10, n. 2, p. 1-27, 2000. Disponível em: <<http://www.ies.ufpb.br/ojs/index.php/ies/article/view/326>>. Acesso em: 26 abr. 2014.

VIDAL, A. A conservação e a preservação de documentos digitais: um desafio na era da sociedade de informação. *Revista da Faculdade de Ciências Humanas e Sociais*, Porto, v. 7, p. 144-154, 2010. Disponível em: <<http://hdl.handle.net/10284/2809>>. Acesso em: 9 abr. 2014