DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646346

ARTIGO

GESTÃO DE REPOSITORIOS DE PRESERVAÇÃO DIGITAL

MANAGEMENT OF DIGITAL PRESERVATION REPOSITORIES

GESTIÓN DE REPOSITORIOS DE PRESERVACION DIGITAL

Miguel Ángel Márdero Arellano[1]
Alexandre Faria de Oliveira[2]

RESUMO

Experiências internacionais de implementação de práticas de preservação digital em repositórios são o objeto deste trabalho. Com base num levantamento bibliográfico sobre o inicio das práticas de preservação em repositórios digitais, foram identificados aspectos ainda relevantes para os gestores desses repositórios. A maioria dos repositórios citados registrava a dupla função de acesso e preservação, mas, poucos podem ser considerados "arquivos obscuros" (*dark archives*), usados apenas para fins de preservação. A aplicação de padrões de preservação digital mostrou que apenas instituições de grande porte possuíam definições detalhadas do que podia ser depositado e o uso que podia ser feitos de materiais armazenados. Os gestores desses repositórios tinham algum tipo de orçamento operacional para realizar atividades de preservação. A maioria dos repositórios citados na bibliografia usava uma combinação de ferramentas comerciais e de software livre. Como conclusão, os registros analisados reforçam a necessidade ainda hoje de aplicação de mais de uma estratégia de preservação digital, do uso do modelo de referência OAIS e de auditorias oficiais no desenho de um repositório de preservação, para manter assim, a flexibilidade na integração de funções e serviços que vão além do repositório.

PALAVRAS-CHAVE: Preservação Digital. Softwares e estratégias de preservação digital. Práticas de preservação digital em repositórios em bibliotecas.

ABSTRACT

The object of this work are the international experiences of implementation of digital preservation practices in repositories. It based on bibliographical survey about the beginning of digital preservation practices in digital repositories, identifying important aspects of how to manage the practices of digital preservation on repositories. Most repositories analyzed showed a double function of access and preservation, but few could be considered to be "dark archives" used only for preservation matters. The application of digital preservation standards showed that large institutions possessed detailed definitions of what stored materials could be filed and used. Repositories managers had some kind of operating budget to carry on preservation activities. Most repositories cited in the bibliography used a combination of commercial tools and free software. As a conclusion, the records analyzed reinforce the need today of the application of digital preservation strategies using the OAIS Reference Model and official audit certification actions on the design of digital preservation repositories, to keep the integration flexibility of functions and services that go beyond the repository.

KEYWORDS: Digital Preservation. Softwares and digital preservation strategies. Digital preservation practices on repositories and libraries.

RESUMEN

Experiencias internacionales de implementación de prácticas de preservación digital en repositorios en bibliotecas, museos y archivos son el objeto de este trabajo. Basado en un levantamiento bibliográfico sobre el

[1]Doutor e mestre em Ciências da Informação (UnB). Tecnologista do IBCT e Coordenador da Rede CARINIANA. Brasília, DF. Email: miguel@ibict.br. ORCID: http://orcid.org/0000-0001-5306-919X.

[2] Pós-Graduado em Sistemas Orientado a Objetos PUC-Brasília. Brasília, DF. Servidor no IBICT. Coordenador de soluções tecnológicas na Rede Cariniana. Email: alexandreoliveira@ibict.br. ORCID: http://orcid.org/0000-0003-0470-4972.

Submetido em: 08/08/2016 – **Aceito em**: 12/09/2016

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646346

inicio de las prácticas de preservación en repositorios digitales, fueron identificados aspectos importantes para los administradores de esos repositorios. La mayoría de los repositorios registraban la doble función de acceso y preservación, pero pocos podían ser considerados “archivos obscuros” (*dark archives*) usados solo para fines de preservación. La aplicación de padrones de preservación digital mostrou que solo instituciones de gran tamaño poseían definiciones detalladas de lo que podía ser depositado o el uso que podía ser hecho de materiales almacenados. Los administradores de esos repositorios tenían algún tipo de subsidio operacional para realizar actividades de preservación. La mayoría de los repositorios citados en la bibliografía analizada usaba una combinación de herramientas comerciales y de software libre. Como conclusión, los registros analizados refuerzan la necesidad de que aún hoy se aplique más de una estrategia de preservación digital, del uso del modelo de referencia OAIS y de las auditorias oficiales en la estructuración de un repositorio de preservación, para mantener de esa forma, la flexibilidad en la integración de funciones y servicios que van más allá de l repositorio.

PALABRAS CLAVE: Preservación Digital. Softwares y estratégias de preservación digital. Prácticas de preservación digital en repositórios y bibliotecas.

1 OS REPOSITÓRIOS E AS FERRAMENTAS DE PRESERVAÇÃO DIGITAL

Este trabalho esta baseado em uma revisão da literatura sobre preservação digital, tendo como foco as primeiras experiências de gestão das estratégias de preservação digital em repositórios. Foram algumas dessas iniciativas que ajudaram a desenvolver os padrões internacionais da preservação atuais. A análise dos dados coletados de textos publicados nas primeiras duas décadas da Internet, visa apontar de forma descritiva o uso de ferramentas tecnológicas e sua normalização em torno dos conceitos e relações detalhadas no modelo de referência OAIS.

Na era dos repositórios digitais a preservação digital tem sido relacionada com o acesso livre, com os repositórios institucionais, com os sistemas de armazenamento e de backup, sendo que ela, em primeiro lugar, é um problema técnico relacionado com todas as atividades de manutenção e de cuidados dos objetos digitais.

Como respostas a esse desafio, surgem no final do século XX estratégias de preservação digital que procuram incorporar todos os aspectos relacionados ao problema tecnológico: custos, legislação, gestão, acesso, políticas e critérios. São formas de reunir soluções parciais ante um problema complexo no qual estão envolvidos, entre outros itens, a migração, emulação, arqueologia digital, criptografia, metadados, formatos-padrão e software livre.

Com o desenvolvimento de pacotes de software, pode-se afirmar que a área de preservação digital está chegando à maturidade. Algumas dessas soluções saíram da área da ciência da informação, como ferramentas e serviços para bibliotecas e arquivos. Entre os primeiros sistemas citados estavam, o LOCKSS (Lots of Copies Keep Stuff Safe), o PANDAS, o OCLC Digital Archive, e o DIAS (Digital Information Archive System) da IBM. Todos eles procediam de organizações de diferentes perfis: provedores de serviço para biblioteca, grupos de pesquisa de bibliotecas especializadas, equipes de repositórios

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646346

universitários e bibliotecas nacionais trabalhando em parceria com empresas terceirizadas, mostrando a necessidade de integração entre todos os interessados no assunto.

As ferramentas para repositórios institucionais não eram originalmente destinadas a projetos de preservação digital. A motivação primordial não diz respeito à garantia da longevidade dos conteúdos digitais, mesmo que elas acompanhem o desenvolvimento nas áreas de preservação em longo prazo de conteúdos digitais. O conteúdo dos repositórios institucionais pode ser preservado, mesmo que seu objetivo não seja a preservação, e principalmente, pois o conteúdo de acesso aberto está em risco de não ser preservado devido ao fato dele não estar sendo totalmente disponibilizado sem restrições.

O modelo de preservação digital nas bibliotecas e repositórios digitais enfoca a necessidade de futuros usuários disporem de materiais autênticos e certificados por instituições reconhecidas. A descrição em metadados de todos os detalhes que expressem a história de criação de um objeto digital está sendo considerada uma metodologia que pode garantir a autenticidade de um registro eletrônico.

Pesquisadores de vários projetos que envolvem *softwares* para repositórios digitais asseguram que os repositórios institucionais são os lugares adequados para testar e formular as metodologias e políticas a serem adotadas pelos provedores de informação científica. O argumento no qual eles se baseiam é que os repositórios para preservação de objetos digitais devem estar localizados em instituições confiáveis e capazes de armazenar, migrar e dar acesso a coleções digitais (OCLC/RLG, 2002).

Os responsáveis pelos acervos digitais confiam nos repositórios institucionais para poder preservar e dar acesso a material não apenas publicado em periódicos avaliados pelos pares, mas também à literatura cinzenta, como teses, relatórios, documentos governamentais e, ainda, materiais suplementares, conjuntos de dados, imagens, visualizações e simulações, comunicação informal como *e-mails*, *blogs*, *podcasts*, *websites*, *wikis* e apresentações.

Os repositórios digitais podem integrar ferramentas de preservação digital, ou aquelas que sejam equivalentes nessa funcionalidade. Uma instituição pode também optar por definir um tipo de *workflow* que integre ferramentas em determinados momentos do processo. Foram identificada algumas das mais citadas na literatura da área, menciondas nas páginas da Rede Cariniana do Ibict:

a) as que geram e capturam metadados: FIDO, Metadata Extraction Tool, NARA File Analyzer and Metadata Harvester, Tika, NLNZ metadata extractor, Exiftool;

b) as que identificam e avaliam os formatos de arquivo: TrID - File Identifier , File Information Tool Set (FITS), JHOVE e DROID;

c) as que padronizam arquivos para formatos preserváveis ou formatos abertos para a preservação: XENA da National Archives of Australia, OpenDataForge;

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646346

d) as que comprimem e descomprimem os arquivos: Winzip, Filezila, 7zip, winrar;

e) as que geram pacotes para envio: BagIt Transfer Utilities, Bagger, Manifest Maker;

f) As realizam analise forense: BitCurator, Autopsy The Sleuth Kit, Encase Forensic, FTK (Forensic Toolkit);

g) as que realizam captura web: WebCopier, HTTrack, Heritrix, Web Curator Tool (WCT), Archive-it;

h) as que realizam captura de base de dados: SIARD;

i) as que realizam o planejamento e o gerenciando dos arquivos de preservação: Data Asset Framework, DMPTool, DPBCT (Digital Preservation Business Case Toolkit);

j) as que realizam auditoria em preservação digital: NDSA Levels of Digital Preservation, Nestor 2;

k) as que conseguem executar funções hash criptográficas: HashX, md5summer, fsum, HashMyFiles, Fixity, md5sum, Checksum Checker;

l) as que executam o controle de integridade de armazenamento: ACE (Audit Control Environment), SAFE Archive Audit System;

m) as que consolidam a migração de formatos: FF mpeg, ImageMagick, XnView, IrfanView;

n) as que converte arquivos de dados em arquivos preserváveis no formato XML: MIXED (Migration to Intermediate XML for Electronic Data);

o) as que realizam backup: Acronis True Image, SyncBack, Cobian Backup, Teracopy;

p) as que fazem armazenamento em nuvem: Amazon S3, Amazon Glacier, Dropbox, Windows Azure, RackSpace;

q) as que são sistemas de preservação digital: Digital Preservation Software Platform (DPSP), Archivematica, LOCKSS, DAITTS, RODA, Preservica, Duracloud;

r) as que realizam a emulação: Virtual Box, VMware, Emulation Framework (EF).

Esses são alguns exemplos de ferramentas e soluções tecnológicas que estão sendo testados e que são considerados padrões de preservação digital pela comunidade científica, mas os pesquisadores estão ainda distantes de uma automação completa em sistemas de preservação digital.

A partir desse princípio, pode-se observar que essas ferramentas contemplam todo um ciclo de vida de preservação digital, durante o qual as organizações responsáveis pela gestão de conteúdo patrimonial digital a longo prazo devem demostrar a aplicação de boas práticas que envolvam um conjunto de padrões de preservação digital. Esses padrões vêm surgindo desde 1996 com o relatório de Preservação de Informação Digital que foi lançado e constitui um guia para a comunidade da preservação digital para as próximas décadas (COMMISSION ON PRESERVATION AND ACCESS AND THE RESEARCH LIBRARIES GROUP, 1996)). Atualmente os padrões de Preservação Digital incluem:

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646346

- TDR – Trusted Digital Repository Checklist (ISO 16363);
- OAIS - Open Archival Information System Reference Model;
- PAIMAS – Producer-Archives Interface Methodology Abstract Standard;
- NISO – Z39.87 Data Dictionary – Technical Metadata for Digital Still Images;
- PREMIS – Preservation Metadata Implementation Strategies;
- BRTF-SPDA – Blue Ribbon Task Force on Sustainable Digital Preservation and Access;
- TRAC – Trustworthy Repositories Audit & Certification.

Internacionalmente, as instituições conseguem o status de repositório digital confiável (TDR) quando demostram concordância com esses padrões e normas. Um repositório digital confiável deve seguir: princípios, políticas, fluxos e prevenção. O resultado da observância desses padrões produz não apenas um plano de preservação digital, mas, uma documentação substancial com dados para curadoria, que sustentem um planejamento de preservação digital adequado para a organização.

Os aspectos organizacionais da preservação digital são mais desafiadores que os assuntos técnicos, por isso muitos membros da comunidade têm se centrado nesses aspectos. Por isso, para dar inicio à prática da preservação digital projetos introdutórios estão sendo construídos ao redor da implementação de uma ou mais ferramentas que permitem a gestão dos conteúdos digitais em diversos níveis do seu ciclo de vida, como na aquisição, no processamento ou no acesso.

Quando um programa de preservação digital madurece, o foco precisa mudar para o tipo de atividades de planejamento da preservação digital que as ferramentas de gestão da preservação digital observam.

As características principais dos repositórios para objetos digitais foram mencionadas em 2002 por Stewart Granger como aquilo que formaria a “infra-estrutura profunda”, relacionada com toda a parte organizacional, aspectos legais e culturais, assim como toda a parte tecnológica. Para ele, o problema principal estaria em que essa infra-estrutura realmente responda às necessidades da instituição e de seus usuários, e não apenas as das empresas que comercializam essas tecnologias. Para que isso seja possível, ele mencionou a necessidade de que existam repositórios certificados, assim como mecanismos de colaboração que facilitem o intercâmbio entre a comunidade envolvida no tratamento de materiais digitais.

Como produtoras de pesquisas científicas, as instituições acadêmicas estão interessadas na captura, disseminação e preservação da produção intelectual de seus próprios membros. Tradicionalmente, as editoras e bibliotecas têm tido o papel complementar de facilitar a publicação e preservação da produção científica. Nas últimas décadas, as mudanças

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646346

tecnológicas e de mercado têm-se acelerado, motivadas em parte pelo volume crescente da publicação de resultados de pesquisas.

Essa mudança de papéis está afetando a relação simbiótica entre editoras e bibliotecas. A combinação de uma rede quase ubíqua com o aumento crescente dos preços nos modelos tradicionais de publicação prepara o cenário para novas expectativas. Entre elas está o uso de repositórios institucionais para prover aos professores novas formas de criar e preservar objetos de aprendizagem, tais como ilustrações, visualizações, modelos e vídeos.

Peters (2002) mencionou que o papel das instituições de ensino superior e de pesquisa envolvidas em projetos de preservação digital seria de detentoras de repositórios digitais, que permitissem controle da autoria e a cobertura de elevado número de áreas de conhecimento.

Outros autores, como Messerschmitt (2003) e Hitchcock (2003), também afirmaram que os repositórios institucionais poderiam ser viabilizados mais facilmente através de sistemas distribuídos de preservação digital. Neles, os repositórios centralizados e mais bem organizados compartilhariam suas metodologias com os outros repositórios temáticos e institucionais. Para Messerschmitt, o papel das bibliotecas como curadoras da informação digital é fundamental, uma função que elas têm realizado por séculos.

Como desde seu surgimento a maioria dos repositórios estava em instituições acadêmicas, bibliotecas universitárias começaram também a desenvolvê-los. As bibliotecas se uniram às agências de governo, sociedades históricas, museus e outras instituições culturais para estabelecer repositórios colaborativos ou temáticos. Nesse intuito, vários repositórios começaram a ser desenvolvidos em consórcio, uma vez que nem todas as instituições acadêmicas precisam ou desejam tê-los.

Os repositórios institucionais estendem significativamente o papel das bibliotecas, representando um compromisso sério e de longa duração, com muitos benefícios. Professores e cientistas que começam a usar repositórios institucionais para publicar e preservar seus trabalhos confiam na integridade, conhecimento e competência daqueles que gerenciam esses repositórios.

A experiência acumulada de adoção de padrões e protocolos de comunicação por parte dos profissionais da informação é fundamental na construção de repositórios institucionais confiáveis. Eles podem definir um conjunto mínimo de critérios para o arquivamento de informações científicas produzidas em uma instituição. Esse conjunto de critérios para o desenvolvimento de um repositório de preservação deve estar influenciado pelo modelo OAIS e pelo princípio arquivístico de custódia responsável.

O contexto dos repositórios pode ser caracterizado pelos seguintes atributos:

a) repositórios que armazenam arquivos que nasceram digitais, sem análogos em papel;

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646346

b) repositórios que possuem atributos de confiabilidade e disponibilidade relevantes para a comunidade científica;

c) repositórios de acesso contínuo, atributo digital que garante a possibilidade de citar, descobrir, entregar e usar o recurso depois de sua criação e depósito no repositório. (MÁRDERO ARELLANO, 2008)

Esses atributos devem ser garantidos permanentemente para prevenir as falhas dos formatos e controlar os efeitos das mudanças tecnológicas. Também, os processos de preservação digital devem estar apoiados em políticas bem definidas, na sua organização e nas estratégias adotadas.

2 OS REPOSITÓRIOS E O MODELO DE REFERÊNCIA OAIS

O modelo OAIS fornece as especificações de um repositório e estabelece responsabilidades que uma organização deve distribuir para operar como um arquivo de acesso livre. O modelo de referência define um sistema de informação para arquivamento aberto como aquele composto por uma organização de pessoas e sistemas que aceitam a responsabilidade de preservar informação e sua disponibilização para uma comunidade específica. A aplicação dos princípios do modelo OAIS, e particularmente, a implementação de um arquivo de acesso livre em concordância com os modelos de funcionalidade e estrutura da informação do OAIS é o pré-requisito chave para estabelecer repositórios confiáveis e garantir a preservação de longo prazo dos seus atributos digitais.

Para garantir a confiança nos repositórios digitais de acesso livre por parte dos autores e dos usuários, eles devem manter algumas propriedades de preservação digital mínimas:

a) autenticidade: a certeza de que um componente digital foi criado pela pessoa que afirma tê-lo feito; a autenticidade permite ter a certeza de que o criador do objeto digital não pode negar que foi ele quem o criou. As assinaturas digitais e as marcas d'água digitais são técnicas que garantem a autenticidade dos objetos digitais;

b) integridade: a habilidade de manter os dados completos e corretos, prevenindo mudanças acidentais ou maliciosas (corrupção dos dados). Entregando e salvando um *bit/byte checksum*, como o MD5 faz, constitui uma técnica básica para detectar se qualquer modificação produziu algum efeito nos objetos digitais depois de ter sido inserido no repositório;

c) confiança e disponibilidade: a confiabilidade está relacionada à habilidade dos componentes de *hardware* e *software* funcionarem de acordo com suas especificações sem erros ou defeitos. Disponibilidade é a porcentagem do tempo que

> o sistema está regularmente em funcionamento, em relação com o tempo total que ele deve operar. Algumas técnicas usadas para garantir altas porcentagens de confiabilidade e disponibilidade são *backups*, *softwares* de antivírus, *firewalls*, *operating system patches*, atualizações de aplicações de *software*, componentes de *hardware* de redundância e tolerância de falhas;
>
> d) capacidade de reuso: habilidade de acessar um recurso digital pelo tempo que a instituição decida manter o repositório. Os objetos digitais científicos e acadêmicos que possuem valor por longo período de tempo devem ser recuperados apropriadamente e reusados ao menos por um período longo de tempo (uso de identificadores permanentes e mantendo formatos e mídias). A preservação digital deve incluir todos os componentes de dados da infra-estrutura dos repositórios digitais de acesso livre, não apenas os objetos digitais, mas também seus metadados e identificadores. (MÁRDERO ARELLANO, 2008)

A confiabilidade vem diretamente do principio arquivístico de custódia responsável. Como observa Thomaz (2007), a confiança nos repositórios digitais se desenvolve em no mínimo três níveis:

1. A confiança de que os produtores estão enviando as informações corretas;
2. a confiança de que os consumidores estão recebendo as informações corretas; e
3. a confiança de que os fornecedores estão prestando serviços adequados.

A primeira lista de atributos e responsabilidades de repositórios digitais confiáveis foi estabelecida pela Research Libraries Group (RLG) e o Online Computer Library Center (OCLC) no seu relatório publicado em 2002. Esse conjunto de atributos está influenciado pelo modelo OAIS e estabelece que essa obediência ao modelo deva ser considerada o primeiro critério a ser observado por um repositório confiável. Para esse grupo, os repositórios confiáveis devem incluir também, atributos que possam dar suporte a sistemas de segurança e aos procedimentos adequados e significativos. Eles apontam que todo repositório confiável deve incluir atributos que sustentem os seguintes aspectos:

a) responsabilidade administrativa;
b) viabilidade organizacional;
c) sustentabilidade financeira;
d) adequabilidade tecnológica e procedimental;
e) sistema de segurança;
f) responsabilidade de procedimentos (certificação).

O relatório da RLG e da OCLC codifica esses atributos no conjunto de responsabilidades baseado no modelo OAIS/SAAI, manifestado suas responsabilidades de

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646346

custódia e definindo, assim, o que pode ser considerado como uma lista de requisitos funcionais (Figura 1).

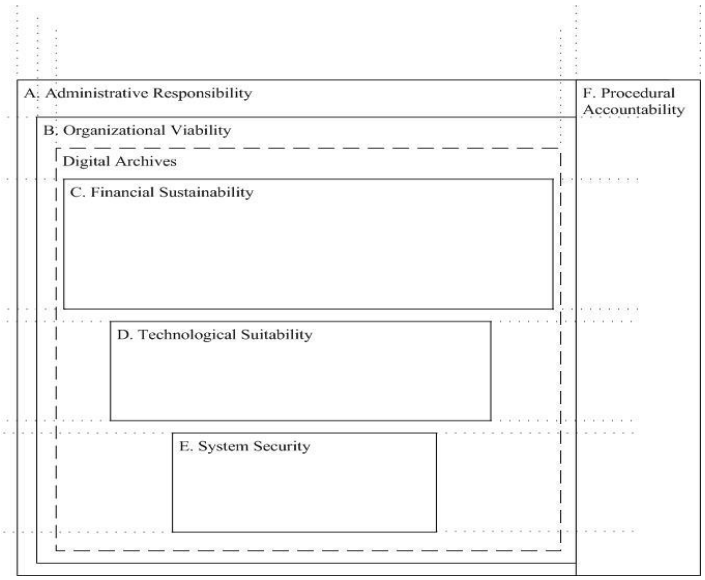

Figura 1. Modelo de *Trusted Digital Repository* (TDR)
Fonte: RLG/OCLC (2002)

O modelo *Trusted Digital Repository* representa a primeira forma de expressar a infra-estrutura organizacional da preservação digital. Os atributos do TDR converteram-se em padrões para a comunidade da preservação digital, pois antes deles não existia uma expressão formal do contexto organizacional da preservação digital.

Uma nova versão do modelo TDR surgiu do projeto de preservação digital da Cornell University, sugerindo que o modelo a ser seguido deve sair do entendimento de dois documentos-chave: o que propõe os atributos de um repositório digital confiável (TDR) da RLG/OCLC na implementação tecnológica, e da proposta do modelo de referência OAIS/SAAI para a estrutura do contexto organizacional. A representação dessa integração de propostas está no diagrama elaborado por Nancy Y. McGovern (2007) como mostra a Figura 2.

Usando o diagrama desenvolvido pela University of Cornell, McGovern aborda a questão da fronteira chamada de *Digital Archives Border* entre os primeiros dois atributos e o restante, para esclarecer sua importância naqueles casos em que uma instituição mantém mais de um repositório. Os primeiros dois atributos se aplicariam a todos os repositórios na instituição, já que mais de uma organização pode chegar a gerenciar apenas um repositório (como no caso de um consórcio).

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646346

Figura 2. Diagrama da integração dos dois pilares da preservação digital baseado no modelo TDR da RLG/OCLC
Fonte: McGovern (2007)

Desde 2003 a Cornell University vem criando parcerias e repassando para a comunidade sua experiência com a aplicação desses princípios, colocando tutoriais *on-line* e cursos para gestores de projetos de preservação digital. A Cornell University foi uma das primeiras instituições de ensino que utilizaram as duas grandes propostas para repositórios confiáveis; com elas foi estabelecida sua política de preservação digital e iniciados alguns dos seus projetos (CORNELL UNIVERSITY LIBRARY, 2004). Também nessa universidade o trabalho seguiu alguns princípios norteadores:

a) focalizar as ações na redução de riscos;
b) proteger as versões arquiváveis;
c) entender os componentes dos objetos digitais;
d) reconhecer algum tipo de perda como inevitável;
e) estabelecer um ciclo de planejamento e manutenção.

Após a definição desses princípios, foi detectada a necessidade do trabalho colaborativo, ou seja, uma única instituição não conseguiria atender todas suas necessidades de preservação. Todos os projetos e iniciativas mapeados buscam a realização da sua intenção original que é comum para todos eles: preservar a produção intelectual e cultural das instituições e garantir o acesso permanente à informação.

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646346

A possibilidade de se criar um modelo genérico de serviços de preservação digital para repositórios institucionais usando o modelo de referência OAIS é uma das necessidades nas iniciativas públicas de projetos e programas colaborativos. Alguns dos projetos começam a elaborar guias e manuais para auxiliar no processo de inserção de dados (*ingest process*), a estimular o depósito de arquivos em formatos-padrão para diminuir custos operacionais de longo prazo e a recomendar melhores práticas a serem implementadas.

Segundo Rosenthald *et al* (2005), a primeira revisão crítica do modelo de referência OAIS foi realizada pelo projeto *Securing a Hybrid Environment for Research Preservation and Access* (SHERPA) projeto encerrado em 2009. Usando o modelo para identificar os direitos e as responsabilidades executadas pelo repositório institucional, foram delineados os requisitos mandatórios e as entidades funcionais dentro do projeto de serviço de preservação distribuída do SHERPA (Figura 3).

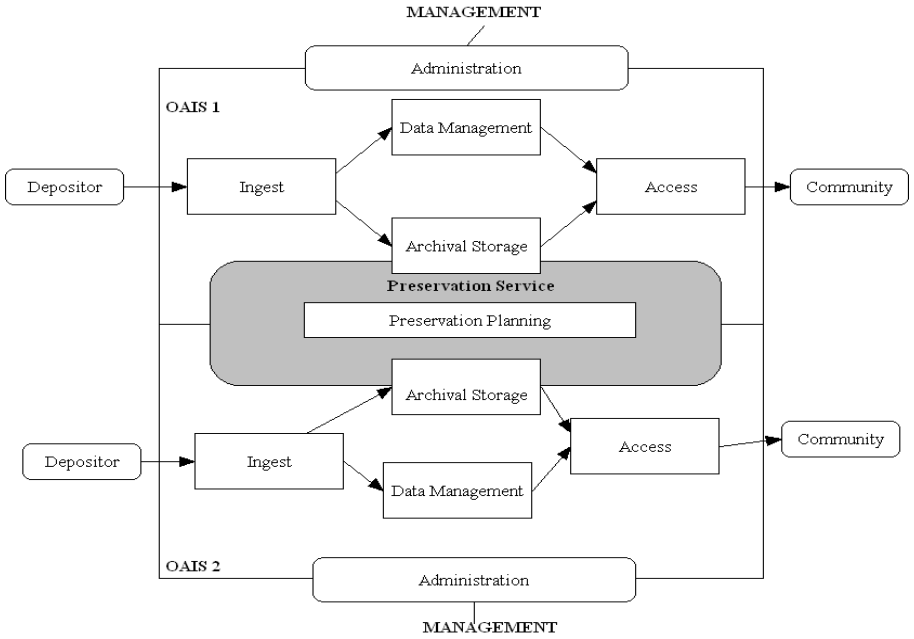

Figura 3. Modelo funcional OAIS no projeto SHERPA
Fonte: Rosenthal *et al.* (2005)

Projetando já um possível processo de auditoria para atribuir/certificar a concordância com o modelo OAIS, as primeiras ações desenvolvidas pelos responsáveis pelo projeto foram:

a) mapear as seis entidades de um repositório que esteja em concordância com o OAIS (inserção, armazenamento arquivístico, administração, gerenciamento de dados e acesso) dentro de uma estrutura existente;
b) garantir que a terminologia de um domínio específico pode ser mapeada por um equivalente OAIS.

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646346

O projeto SHERPA mostrou que a auditoria pode começar no processo de inserção dos dados, já que os sistemas de repositórios digitais podem ser divididos em dois grupos, aqueles em que o autor ou editor deposita diretamente o conteúdo (ex.: DSpace), e os que usam algum tipo de coleta desde as páginas a repositórios na *web* (ex. LOCKSS). Segundo Rosenthal, os dois processos não eram imunes a algum tipo de ameaça e a auditoria pode confirmar a autenticidade do conteúdo inserido (ROSENTHAL *et al*., 2005).

3 ATRIBUTOS DE REPOSITÓRIOS DIGITAIS CONFIÁVEIS

Os atributos e responsabilidade dos repositórios digitais em uso atualmente no exterior estão sendo definidos dentro do contexto dos desafios relacionados com a habilidade das instituições de integrar o gerenciamento de materiais digitais na sua estrutura organizacional. Outra característica desses centros é ter entre seus objetivos a pesquisa e o desenvolvimento das melhores práticas para maximizar os benefícios das novas tecnologias.

O reconhecimento da importância dos atributos de preservação digital e as políticas e procedimentos para sua aplicação constituem a primeira linha de ação no estabelecimento de sistemas de gerenciamento de informação digital.

Desde suas primeiras implementações, os atributos de preservação digital podem ser acompanhados já no processo de escolha das ferramentas e formação dos repositórios digitais (CatalysIT, 2006; WHEATLEY, 2003). Com base na aplicação dos atributos TDR e do modelo OAIS, os gestores desses sistemas de gerenciamento de informação digital podem propor algumas ações técnicas e organizacionais;

1) avaliar o funcionamento das primeiras migrações para novos formatos;
2) analisar os indicadores de desempenho dos repositórios e ferramentas adotadas;
3) definir políticas de seleção e depósito e retenção de documentos;
4) definir os perfis dos responsáveis e tipo de aperfeiçoamento para cumprir eficientemente as tarefas designadas.

Para efetivar as duas primeiras ações, alguns atributos técnicos de preservação digital têm sido usados, tais como:

1) segurança;
2) interoperabilidade;
3) qualidade e capacidade de configuração das ferramentas no *workflow;*
4) internacionalização – interfaces multilíngües;
5) licença para *software* livre.

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646346

Para as duas últimas ações, duas responsabilidades orientadas para os aspectos organizacionais são as mais indicadas: facilitar a configuração segundo o tipo de usuário e ter o suporte da comunidade.

4 CERTIFICAÇÃO DE REPOSITÓRIOS DIGITAIS

Em países como os Estados Unidos, Inglaterra e Alemanha já estão em funcionamento iniciativas de auditoria de arquivos digitais. Instituições nos Estados Unidos como o Research Library Group (RLG) e o National Archives and Records Administration (NARA) que criaram um grupo de trabalho sobre certificação de repositórios digitais e na Inglaterra como o Center for Research Libraries (CRL) com seu projeto de auditoria e certificação de arquivos digitais, produziram um conjunto de princípios de auditoria em comum. São iniciativas que estão focalizando os benefícios e as ferramentas necessárias para a autovalorização e auditorias terceirizadas.

Sobre a capacidade de auditoria e certificação, existem iniciativas em funcionamento que vêm proporcionando ferramentas para auto-avaliação, mas as organizações precisam de meios para participar de auditorias externas. Processos de certificação podem criar maior padronização e credibilidade dos arquivos digitais que vão ao encontro das necessidades das bibliotecas e seus usuários. Os exemplos locais e as lições aprendidas estão contribuindo nesse sentido.

Com o reconhecimento da importância das políticas de preservação, ficou implícito o papel que elas têm dentro dos requisitos de evidência necessários para criar mecanismos de auditoria e certificação dos documentos e dos repositórios digitais. Em 2006 a RLG/NARA publicou o documento *Audit Checklist for Certifying Digital Repositories*, e no ano seguinte o Center for Research Libraries CRL lançou o *Trustworthy Repositories Audit & Certification: Criteria and Checklist (*convertida em norma ISO em 2012), demostrando que as auditorias identificam realmente os pontos fortes e fracos nos programas de preservação digital, e como elas podem ajudar a definir planos de desenvolvimento que gradativamente cumpram com o conjunto de critérios definidos para os repositórios digitais confiáveis (McGOVERN, 2007).

A *checklist* da RLG/NARA define o conjunto de políticas gerenciais que são organizadas como critérios de preservação digital pela instituição, as funções do repositório, os processos e procedimentos, a comunidade alvo a usabilidade da informação e as tecnologias e a infra-estrutura técnica (RLG/ NARA, 2006).

A adoção de ferramentas de preservação digital que podem ser personalizadas, intercambiáveis e adaptadas aos *workflows* nos repositórios digitais está levando as

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646346

instituições a abrir espaços para integrar requisitos de auditoria e de medição, dos pontos fracos e fortes dos programas de preservação digital.

Entre as conclusões a que chegaram os pesquisadores do RLG-NARA, no seu documento *Criteria for Measuring Audit Checklist for Certifying Digital Repositories,* está a de que as ferramentas usadas para fazer auditoria dos repositórios precisam ser desenvolvidas pelos próprios executores dos projetos de preservação digital.

O estabelecimento de programas de certificação e de critérios para serem usados em auditorias tem sido uma necessidade identificada na comunidade da ciência da informação que lida com repositórios institucionais e que precisa de modelos de certificação dessas ferramentas (LYNCH 2003).

Algumas iniciativas que usam ferramentas como JHOVE, DROID e XENA apresentaram, além da funcionalidade de preservar os objetos digitais na sua integração dessas ferramentas aos *softwares* de repositórios digitais, elementos que propiciam a comparação dos limites e capacidades das organizações no cumprimento dos requisitos de preservação apontados pelos modelos TDR e OAIS.

Projetos como o *Network of Expertise in Long-Term Storage of Digital Resources* (Nestor) da University Library of Humboldt-University Berlin, que desenvolveu seu "Catálogo de Critérios para Repositórios Digitais" são exemplos do crescente movimento para o desenvolvimento de parâmetros para medir a qualidade o confiabilidade de um repositório. O grupo de trabalho do Nestor tem entre seus objetivos "[...] formular critérios para repositórios digitais confiáveis e recomendações para procedimentos de certificação de repositórios digitais [...]" (DOBRATZ; SCHOGER; STRATHMANN, 2007 s/p.).

A complexidade dos sistemas de preservação digital em repositórios confiáveis foi apontada pelo grupo de trabalho do projeto Nestor em 2004; ele destacou a importância de abordar todo o processo gerencial de informação digital das organizações onde se deseja que a confiabilidade seja confirmada (NESTOR, 2006). A partir daí, todo o ambiente da preservação digital começou a ser tomado em consideração, assim como a recomendação de que os resultados da auditoria fossem comunicados de forma transparente ao público para gerar mais confiança.

Propostas como a do catálogo do projeto Nestor foram usadas como instrumentos para auto-avaliação das etapas de desenvolvimento de projetos de repositórios digitais confiáveis. Sua aplicabilidade internacional tem suas limitações por razões geopolíticas, mas é levada em consideração internacionalmente para o estabelecimento de um processo formal de certificação ISO.

Com o lançamento do "*checklist" da CRL Criteria for Measuring Trustworthiness of Digital Repositories and Archives: an Audit & Certification Checklist*, representantes de

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646346

projetos de vários países construíram uma colaboração formal para auditoria e certificação de repositórios de preservação digital (CENTER FOR RESEARCH LIBRARIES, 2007).

Tanto o Grupo de Trabalho do projeto NESTOR na Alemanha, quanto o DCC na Inglaterra e o CRL nos Estados Unidos coincidiram nos princípios básicos para a aplicação de critérios institucionais de preservação digital, como segue:

1) Documentação (evidência): as metas, os conceitos, especificações e implementações de preservação digital de um repositório deverão ser documentadas adequadamente. A avaliação inicial do repositório como um todo baseada na documentação pode prevenir erros e implementações inapropriadas.
2) Transparência: a publicação da documentação leva à transparência. Segundo a CRL, "[...] apenas o repositório que expõe seu *design*, especificações, práticas, políticas e procedimentos para análise de riscos pode ser considerado confiável [...]" (CENTER OF RESEARCH LIBRARY, 2007, s/p).
3) Proporcionalidade: nenhum padrão ou norma deve ser tomado como único e absoluto para avaliação de todos os aspectos de um repositório digital, mas ela deve ser realizada baseada nos objetivos e tarefas aplicadas à preservação digital.
4) Mensurabilidade: apenas para alguns casos, devido aos aspectos de temporalidade dos processos de preservação, algumas formas de controle não são viáveis, mas a instituição deverá informar os indicadores para medir o grau de confiabilidade, segundo seu nível de transparência.

As conclusões a que esses projetos chegaram são de que a certificação dos repositórios envolve mais do que a aplicação de critérios resultantes desses princípios básicos. Ela deve prover ferramentas para o planejamento da auto-avaliação e formas de auditorias internas e externas. O contexto de ambas é geopolítico, mas o processo não varia significativamente.

A comunidade de desenvolvedores desses repositórios vem confirmando a importância do processo de auditoria com mais frequência do que o de certificação, pois a auditoria permite que os desenvolvedores dos repositórios analisem e respondam de forma sistemática às carências e acertos na criação dos sistemas. A discussão continua estando centrada na aplicação de procedimentos de auditoria automatizados de contextos específicos da preservação digital, validando políticas locais definidas na aplicação de *softwares* para repositórios (MOORE; SMITH, 2007).

Para os especialistas em preservação digital, um apoio veio ao serem formalizados os primeiros centros de curadoria digital em 2005 (FULTON, BOTICELLI e BRADLEY, 2011), instituições que trabalham nesse sentido são a *Digital Curation Centre* (DCC) na

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646346

Inglaterra e a UNC Chapel Hill nos Estados Unidos. O DCC é um centro que oferece serviços e produtos para a comunidade que trabalha com a curadoria de materiais digitais, oferecendo apoio à comunidade para trabalhar em rede e facilitar formas de acrescentar valor aos conteúdos digitais. Outras instituições estão discutindo a criação de programas de pós-graduação dentro das comunidades de ciência da informação, biblioteconomia, arquivologia e museologia, para capacitar seus estudantes nos trabalhos na área da curadoria da informação digital.

5 CONSIDERAÇÕES FINAIS

A análise na bibliografia sobre os repositórios digitais mostrou que estes continuam definindo os níveis de serviços de preservação que podem oferecer, dependendo do conteúdo e do tipo de objeto. Para alguns conteúdos o repositório pode comprometer-se a chegar no nível da preservação dos *bits*, garantindo retornar ao autor do depósito a sequência de *bits* sob demanda. Para outros registros, alguns sistemas de repositórios podem aplicar estratégias de preservação como a migração ou até trabalhar em parceria com outros repositórios usuários das mesmas ferramentas.

Mas a falta de políticas de preservação na maioria dos projetos de repositórios digitais sugere a carência de conhecimentos técnicos sobre a importância das estratégias de preservação de materiais digitais existentes. A literatura da área mostra que a maioria dos projetos está nos seus primeiros estágios e apenas metade deles relata seguir alguma forma de modelo de negócios e processos de auditoria planejados.

Embora o contexto atual dos trabalhos sobre auditoria e certificação de repositórios de preservação não proporcionem ainda elementos suficientes para perceber os benefícios obtidos na aplicação auditorias de preservação digital, espera-se que a definição de normas de auditoria e certificação como a ISO 16363 seja reconhecida internacionalmente e que sua assimilação geopolítica mantenha as características interoperáveis dos sistemas.

Padrões, normas e práticas de preservação são adotados, quando existe consciência e interesse por parte dos responsáveis pelos acervos em desenvolver recursos que sejam permanentemente acessados. Em nível nacional, a definição das políticas, obrigações e metodologias mais apropriadas para a preservação dos documentos eletrônicos deve levar em consideração a implementação de pacotes de *software* livres a fim de verificar se atendem às necessidades das instituições de ensino e pesquisa e se estão em concordância com os já testados padrões internacionais, que promovem o arquivamento digital da produção científica de longo prazo (UNESCO, 2007).

No Brasil há desconhecimento ou uma pouca relativa experiência com as práticas de preservação digital por parte dos gestores dos repositórios. Os gestores necessitam de apoio tecnológico em todas as fases do ciclo de vida da preservação digital. Cada uma contendo obrigatoriamente o uso de ferramentas juntamente a utilização dos modelos que forneçam as especificações de um repositório e suas responsabilidades.

REFERÊNCIAS

CATALYST. **Technical evaluation of selected open source repository solutions on behalf of CPIT**. Version 1.3 approved. Wellington, New Zealand: [s. n.], 2006.

CENTER FOR RESEARCH LIBRARIES. **Trustworthy Repositories Audit and Certification (TRAC)**: criteria and checklist (TRAC). 2007. Disponível em: <http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91>. Acesso em: 14 mar. 2016.

COMMISSION ON PRESERVATION AND ACCESS AND THE RESEARCH LIBRARIES GROUP. **Preserving digital information**: report of the task force on archiving of digital Information. Washington, 1996. 64 p.

CORNELL UNIVERSITY LIBRARY. **Digital preservation management**: implementing short-term strategies for long-term problems. 2004. Disponível em: <http://www.library.cornell.edu/iris/tutorial/dpm/>. Acesso em: 16 nov. 2015.

DOBRATZ, S.; SCHOGER A.; STRATHMANN, S. The nestor catalogue of criteria for trusted digital repository evaluation and certification. **JoDI**, v. 8, n. 2, 2007. Disponível em: <http://journal.tdl.org/jodi/issue/view/34>. Acesso em: 14 maio 2016.

FULTON, Bruce; BOTTICELLI, Peter; BRADLEY, Jana. Digln: A Hands-on Approach to a Digital Curation Curriculum for Professional Development. **Journal of Education for Library and Information Science**, v.52, n.2, abr. 2011, p. 95-109.

GRANGER, S. Digital preservation and deep infrastructure. **D-Lib Magazine**, v. 8, n. 2, Feb. 2002. Disponível em: <http://www.dlib.org/dlib/february02/granger/02granger.html>. Acesso em: 28 set. 2015.

HITCHCOCK, S. **Metalist of open access e-print archives**: the genesis of institutional archives and independent services. **ARL Bimonthly Report**, n. 227, p. 4-11, Oct. 2003.

LYNCH, Clifford A. **Institutional Repositories**: Essential Infrastructure for Schorlarship in the Digital Age. ARL, no. 226 (February 2003): 1-7. Disponível em: <http://www.arl.org/newsltr/226/ir.html>. Acesso em: 23 de junho 2016.

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646346

MÁRDERO ARELLANO, Miguel A. **Critérios para a preservação digital da informação científica.** 2008. 356 f. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2008. Disponível em: <http://repositorio.bce.unb.br/bitstream/10482/1518/1/2008_MiguelAngelMarderoArellano.pdf>. Acesso em: 04 Jan. 2016

MCGOVERN, N. Y. A digital decade: where have we been and where are we going in digital preservation? **RLG DigiNews**, v. 11, n. 1, Apr. 2007. Disponível em: <http://digitalarchive.oclc.org/da/ViewObject.jsp?objid=0000070519&reqid=8514>. Acesso em: 28 ago. 2016.

MESSERSCHMITT, D. **Opportunities for Libraries in the NSF Cyberinfrastructure Program**. 2003. Disponível em: <http://www.loc.org/standards/mets/>. Acesso em: 14 out. 2016.

MOORE, R. W.; SMITH, M. Automated validation of trusted digital repository assessment criteria. **JoDI**, v. 8, n. 2, 2007. Disponível em: <http://journal.tdl.org/jodi/issue/view/34>. Acesso em: 21 ago. 2015.

NESTOR WORKING GROUP ON TRUSTED REPOSITORIES CERTIFICATION. **Catalogue of criteria for trusted digital repositories**. 2006. Disponível em: <http://edoc.hu-berlin.de/series/nestor-materialien/8en/PDF/8enpdf>. Acesso em: 6 nov. 2015.

OCLC/RLG Working Group. 2002. "Trusted Digital Repositories: Attributes and Responsabilities". Report by the OCLC/RLG Working Group on Digital Archive Attributes. Montain View, CA.: Research Libraries Group, Inc. Disponível em: <http://www.rlg.org> Acesso em: 22 de maio 2016.

PETERS, T. Digital repositories: individual, discipline-based, institutional, consortia, or national? **Journal of Academic Librarianship**, v. 28, n. 6, p. 414-417, Nov. 2002.

RLG-NARA. Audit checklist for certifying digital repositories. [S. l.]: RLG-NARA Task Force on Digital Repository Certification, 2006. <http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf>. Acesso em: 15 nov. 2015.

RLG/OCLC. **Trusted digital repositories**: attributes and responsibilities. Montain View, Canada: RLG-OCLC Report, 2002. Disponível em: <http://www.rlg.org/longterm/repositories.pdf>. Acesso em: 23 jun. 2016.

ROSENTHAL D. S. H. et. al. Requirements for digital preservation systems: a bottom-up approach. **D-Lib Magazine**, v. 11, n. 11, Nov. 2005. Disponível em: <http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>. Acesso em: 28 jul. 2016.

THOMAZ, K. Repositórios digitais confiáveis e certificação. **Arquivistica.net**, Rio de Janeiro, v. 3, n. 1, jan./jun.2007. Disponível em: <http://www.arquivistica.net/ojs/viewarticle.php?id=118>. Acesso em: 5 mar. 2016.

DOI: http://dx.doi.org/10.20396/rdbci.v14i3.8646346

TDR – Trusted Digital Repository Checklist ISO 16363 / TDR , K. Repositórios digitais confiáveis e certificação. Disponível em: <https://www.crl.edu/archiving-preservation/digital archives/metrics-assessing-and-certifying/iso16363>

UNESCO. **Recomendações sobre Software Livre para Repositório e Sistema de Preservação**. Paris: UNESCO, 2007.

WHEATLEY, P. **A way forward for developments in the digital preservation functions of dspace**: options, issues and recommendations. 2003. Disponível em: <http://dspace.org/news/articles/DpAndDSpace.pdf>. Acesso em: 14 jul. 2016.

ISSN 1678-765X 17>
9 771678 765003